CROSS-CULTURAL DIFFERENCES IN CROSSMODAL CORRESPONDENCES BETWEEN WESTERN AND XHOSA CHILDREN: IMPLICATIONS FOR DESIGN

Corral Cerezo, Víctor

Curs 2014-2015

Director: Salvador Soto-Faraco

GRAU EN ENGINYERIA EN SISTEMES AUDIOVISUALS

upf. Universitat Pompeu Fabra Barcelona
Escola Superior Politècnica

Treball de Fi de Grau

CROSS-CULTURAL DIFFERENCES IN CROSSMODAL CORRESPONDENCES BETWEEN WESTERN AND XHOSA CHILDREN: IMPLICATIONS FOR DESIGN.

Víctor Corral Cerezo

TREBALL FI DE GRAU

GRAU EN ENGINYERIA EN SISTEMES AUDIOVISUALS DE TELECOMUNICACIONS

ESCOLA SUPERIOR POLITÈCNICA UPF

2015

DIRECTOR DEL TREBALL

Salvador Soto-Faraco

Universitat Pompeu Fabra Barcelona

Acknowledgements

First of all I would like to thank my tutor, Salvador Soto-Faraco, for accepting my project idea, guide me through it and for his help and patience during the working period and accepting to help me from the distance while I was living in South Africa.

To all my classmates that made this journey more pleasant and interesting, and helped me in the worst moments, and motivated me for working harder. Especially to Lidia Rodríguez, Pau Fabregat, Fabia Serra, Xavier Ramos and Jordi Romera for all the good times together and the patience they had for having me as a close friend.

To all the teachers that helped me through this studies with their experience and good advices, especially to Jesús Ibáñez, for always being ready to answer my doubts about actual and future studies and for the help he provided me during the realization of this project. Also to Luca Bonatti for his good and needed tips on how to experiment with young children.

To Paloma, for always being there sharing with me all the best and worst moments and supporting me in all the decisions I made.

And finally and most importantly to my parents and sister Aurelio, Montse and Laia for all the effort they put on raising me and being next to me with their support, their confidence and for believing and financing my education.

Abstract

This project addressed cultural differences in perceiving crossmodal correspondences, namely, the mental link that often exists between sensory concepts such as visual height and auditory (tonal) height. We studied western children and children from the Xhosa tribe (South Africa). The goal is to try and understand cultural aspects of crossmodal correspondences and more important for us, to understand how such cultural differences can be used to improve future technological designs, towards generating something similar to a collaborative design.

We present an in-depth review of literature on crossmodal correspondences that grounds the empirical studies (experiments) presented later. These studies are partly based on known basic concepts and partly try to go further in aspects not studied before. An adapted experimental protocol was designed to test a group of spatial crossmodal correspondences we thought would be the most interesting for our goal such as auditory pitch with visual size, lightness and spatial position. We also included specifically technological correspondences, such as scroll up/down and zoom in/out with finger movements and, finally, the known "Takete"/"Maluma" metaphoric correspondence.

The results from the Xhosa children were compared to the ones from western culture children and obtained conclusions to be used to apply on all type of designs that would help the Xhosa community to interact better with applications, technology or others, designed using this concepts, making the whole user experience more attractive and pleasant. These results also speak to the cross-cultural generality of some cross-modal correspondences.

Resum

Aquest projecte ha estat dirigit en veure les diferencies culturals en percebre correspondencies crosmodals, com el link mental que sovint existeix entre conceptes sensorials com l'altura visual i auditiva (tonal). Hem estudiat nens occidentals i de la tribu Xhosa (Àfrica Del Sud). L'objectiu és probar d'entendre els aspectes culturals de les correspondencies crosmodals i més important encara per nosaltres, entendre com les diferències culturals poden ser usades per millorar dissenys tecnològics del futur, generant amb aquestes informacions una cosa similar a un disseny col·laboratiu.

Presentem un un profund estudi de literatura sobre correspondencies crosmodals que serveix com a base pels nostres estudis empirics (experiments) que presentem posteriorment. Aquests estudis estan basats per una banda en conceptes basics ja coneguts i per l'altre en intentar arribar més lluny en aspectes no estudiats anteriorment. Vam crear un protocol experimental adaptat per testejar una serie de correspondencies crosmodals que vam considerar que serien les mes interesants pel nostre objectiu com, to auditiu i mida, lluminància, i posició espacial. També vam incloure correspondències específicament tecnològiques com fer pujar i baixar el contigut d'una pantalla o fer i truere zoom d'una imatge amb els moviments dels dits, finalment, la correspondència metafòrica de "Takete"/"Maluma".

Els resultats dels nens de la tribu Xhosa es van comparar amb els dels nens occidentals i vam obtenir una serie de conclusions usades per aplicar en tot tipus de disenys que afavoririen a la comunitat Xhosa a interactuar millor amb aplicacions, tecnologia o altres, disenyats usant aquest concepte, fent el total de la experiència d'usuari mes atractiva i plaent. Aquests resultats també parlen de la generalitat cross-cultural d'algunes correspondencies crossmodals.

Index

Figure List

"One side effect of today's technologically advanced world is that is is not uncommon to hate the things we interact with. Consider the rage and frustration many people feel when they use computers. In an article on "computer rage", a London newspaper put it this way: "It starts out with a slight annoyance then the hairs of your neck start to prickle and your hands begin to swat. Soon you are bangging your computer or yelling at the screen, and you might well end up belting the person sitting next to you".

(Norman, 2004)

1. INTRODUCTION

The idea expressed in the quote above, by D. Norman illustrates the situation in design for technology back in 2004, when his famous book was published. Even though the deisgn for technology has improved with years, and design in all types of technologies increased in a positive way, we are still far from perfection. Design naturally prioritizes use over beauty, we still put first the usability than the feelings that produces to the user because, in the end, as Donald Norman says, an attractive design is not necessarily the most efficient. And the question is: Could this two concepts go together? Could we design having both ideas in mind and not prioritizing one over the other?

We think that a compromise is possible and we are still far from designing as well as we could, giving the user a perfect sensation and feeling that would make her have a pleasant and attractive interaction with technological devices. There is a good example from the book mentioned above, that makes us think how emotions play a big role in our everyday life interaction with technology; at the beginning of the personal computer era, most of the monitors were in black and white, posteriorly in the early 80's colour screens were introduced, and even though from a cognitive point of view they were not adding any information (anything representable in color was possible to be represented in grayscale) and they were more expensive, people nevertheless started to move from old monochromatic monitors to the new color ones, just because they attached an emotional aspect to them.

Nowadays we see numerous studies trying to understand how the human brain works, how many situations affect perception and cognition and how these aspects are applied into some type of design, for example into choosing the color of the packaging of crisps (Piqueras-Fiszman, & Spence, 2011). One of the phenomena that are related to these ideas are the crossmodal correspondences and we think that understanding and using this mental connections across sensory modalities properly we could get to the point of the better design we are expecting. We can think of crossmodal correspondences as the connection between our senses in perceiving an input of information. How, for example, a higher pitch in audition makes us associate it with a brighter color or a smaller object.

In addition, we think that there is not a universal design, or at least, a design based or generated on the beliefs or likings of a certain group of population that will work for all the other cultural groups. That is why we think it is really important to adapt design to the cultural environment and we can just do this by learning from that environment and the behavior of their population.

One of the bigger discussions in crossmodal correspondences is the doubt of if they are innate or learned (we will talk about this deeper during the project), and the main way to work on the answer of this doubt or question is to study and compare results with children, because this could show if experience plays a big role on crossmodal correspondences or not. That is why in our project we decided to experiment with children from two really different cultures and from each of the cultures with two age groups, to be able to compare results between the cultures but also between the different ages.

In the first section of this report we explain our goals and motivations as well as the structure of our project, to give an idea of the what will be found in the following chapters.

1.1. Objectives

The main goal of this project is to be able to collect a serie of information and theories, to be able to generate in the future a better, more intuitive and pleasant design centered on specific cultures. Our specific way to do this is to understand properly the

crossmodal correspondences and how they change among cultures, more specifically between the western countries (that are the ones mostly represented and studied in most of the researches) and the Xhosa tribe in South Africa.

Some of the typical cross-modal correspondes we, as adults and technology users, give for granted, have not been tested in other cultures. Specially, kids from other cultures who may have rarely shared the same technological experience as western users have. We want to understand how differently the Xhosa kids perceive these correspondences in relation to western kids and how that can affect in the way of interacting with technology. We want to use this information obtained from our study to generate a protocol or a concept for design, to guide future designers. The steps we followed to obtain these objectives are summarized below:

The steps to follow to obtain the different objectives are the following:

- Read and learn the current literature about crossmodal correspondences and the most relevant studies done in this field.
- Study and get immersed in the Xhosa culture while interacting with potential participants of the experiment.
- Make a selection of participants.
- Design the experiment.
- Experiment with the Xhosa kids selected.
- Experiment with the western kids selected.
- Analyze and compare the results.
- Extract the relevant information from the results and generate a theory or guide for future design.

1.2. Motivation

While studying my degree in Engineering of Telecommunications of Audiovisual Systems, I have always been attracted by subjects that have had humans at its focus,

such as “Perception and Cognition”, “Human Computer Interaction”, “Interaction Design”, “Perception Applied on Design” amongst others. All these different subjects have had in common also an important aspect for me, and one of my main interests in life, that is design. I realized that the connection between design and the interaction that humans can do with it is something that really interests me and that puts together the technical part of these specific studies and the artistic part.

Another of my main interests in life is travelling and everything that brings with it, like learning about new cultures and getting involved in the everyday life of the places I visit. By the time I was done with my presential lessons in university I decided to move to Cape Town, South Africa to live for a while and experience a new atmosphere. By that time I started with this final project for my degree, and I thought it would be nice to connect these two of my biggest interests in life, working on a project that would put together design in technology and culture.

The process for defining what the project was going to be about was long and hard, because there were many possibilities that could connect both of my interests and we did not know exactly the situation that I would find in South Africa and the possibilities that I would have in terms of research. That is why I took a long time with a long process of trying to understand firstly the cultural situation of the area I was living and trying to see which aspects would be more interesting to work on the project.

At one point, and after reading a lot about other studies in many different fields, that were having something in common with my interests, we realized that one of the biggest cultural aspects that stand out when you are in South Africa is the many cultures and languages that coexist in the same area and also the big poverty problems that occur all over the country. These two aspects have the consequence that many people in that context have very limited or no access to technological tools that are, otherwise, common in our western environments. That is how we came up with the idea of studying the differences in perception of the crossmodal correspondences, how they could have an impact on interactions with technology, and to understand how this differences could be involved in generate a better design to make this people happier with the technology they could interact with.

1.3. Work Structure

We organized this document the closest way as possible to the way we worked on the project to produce to the reader the same feeling we had while working on it:

The second section of this document is the theorical section of the project, in here we tried inform about all the important theorical information we considered to be known before going over the rest of the sections of the project. Mainly this section is a little abstract of the many concepts and theories we learned while reading about the different subjects contained in this project. This section is devided in different subsections trying to touch some of the most inportant aspects of the theory.

Afterwards on section number 3, we put in context the reader in terms of the cultural atmosphere and background of the group of population we used for our experiment. In this section we sitúate geographicaly this specificl cultural group, we talk about the cultural and economic differences in the área and some other aspects that we think are really relevant to understand why we chose this specific group and location.

In section number 4 we get into our experiment, and we define in different subsections all the details about our experiments, from the initial design till the conclusions obtained from the results. As many other researches we describe with detail the different important aspects of the process, talking about the objectives, the methods, the results and others.

Further on the section number 6 explains our ideas and thoughts of what could be done with the information we obtained from this experiments, and how we think that could be used for fullfilling our main interests of a better and more intuitive technological design.

Finally we use section number 7 to explain the conclusions we obtain from the whole process of this project and how do we think this could be extended and improved with further work done.

2. THEORETICAL FRAMEWORK

2.1. Multisesory Perception

In our everyday life our senses are constantly receiving inputs from what is surrounding us. This information, generally, does not come isolated, but it comes in different sensory channels that complement each other and support a more complex and better perception of the situation, increasing the salience of some events, helping in the resolution of perceptual ambiguities and inducing a unified perception of the surroundings.

Let's imagine a common situation: we are working on a project sitting in front of our table. And we receive a call from a friend. Our telephone is laying on the table a meter away from where we have our computer, as soon as the phone reacts to let us know, that there is a call for us, it does not do it just by ringing, we receive this information by vision, with the phone being lighted up, by audition, through the ring of the phone and by haptics, through the vibration of the phone on the table. In this case we can see easily how our senses get different informations from the same story, binding them for generating a better understanding of it and being able, in this case, to detect the call even if some of our senses are unavailable (we are listening to music over headphones; we are not looking in the direction of the phone...) or simply, to reach the phone faster.

Even though we are able to realize that this is the way we perceive information, our nowadays understanding of how the brain works and processes all this multiple sensory inputs and generates a single perception, that is what we finally receive, is still far from complete. Actually, crossmodal integration is one of the most complicated and complex aspects of the human perception studies. In the following pages of this chapter we are going to go through some relevant points of this subject in relation to our goals.

2.2. Crossmodal Correspondences

How do shapes sound? Ask themselves (Deroy & Auvray, 2013). Which is the color of sweetness? What is the shape of bitterness? (Wan et al., 2014). Even if this questions seem to make no sense, because shapes don't have a related sound, a sweet thing can be

painted in different colors and a bitter aliment can be shaped various ways, there are many studies in cognitive science that prove that exist psychological associations that make this questions have systematic answers, when asked in the right way to people.

A fanmous example of cross-modal correspondence is the case of “Takete” and “Maluma” (Kohler, 1929, 1947). When the participants in this test were asked to associate the sound of each of the above words with the following two shapes in Figure 1, most of them answered in the same way; “Maluma” was the rounded shape and “Takete” was the anglular one. This is interesting, since the observers had never before experience the words nor the shapes in their lives. Why were they so sure of which corresponded to which? Similar to this there are many tests that have shown a common way of processing this type of associations by the participans between different pairs of senses.

Figure 1. “Takete “– “Maluma” Shapes

The past years there have been a big interest in the crossmodal correspondences topic, due to the possible use in applications, the interests in understanding what makes different people from different backgrounds agree in the same, supposed, arbitrary connections and also due to the low cost of the experiments with the ones can be tested. Even if the popularity of these studies has been increasing the past years, and the investigation on multisensory associations has been on for a long time, there are still a lot of aspects to be solved. Where do crossmodal correspondences come from? Are they

innate or learned by experience? Are they shared by all the cultures in the world? And some other questions that we will try to go over during the project and find out as much as we can.

2.2.1. Terminology and Definition

During the pass of the years different terminology has been used to describe the non arbitrary associations between two or more basic stimulus of different senses that seem to be present in humans, some more common than others and that make possible to bind together the different stimulus that we perceive at the same time form the same source: synaesthetic correspondences (Braaten, 1993; Martino & Marks, 2000; Parise & Spence, 2009; Walker et al., 2010), synaesthetic associations (Parise & Spence, 2008), crossmodal equivalences (Lewkowicz & Turkewitz, 1980), crossmodal similarities (Marks, 1987a, 1987b, 1989a, 1989b), and natural crossmodal mappings (Evans & Treisman, 2010).

The various terms listed above have been used for naming similar ideas during the studies on the field, but there are a few differences between them. On the one hand, we use the term *synaesthetic correspondences* and *synaesthetic associations* to talk specifically about the nonredundant associations, for example between size in vision and pitch in audition. That is, between sensory features that are intrinsic of their respective modalities (there is in principle no auditory size, or visual pitch). On the other hand we find the rest of terms used to refer to any kind of different association (including non redundant and redundant). As opposed to non-redundant associations, descrived above, the redundant associacions are those between sensory features that can be perceived through different sensory modalities, for example the size of an object, that can be perceived by vision and haptics. Even if it might look simple and clear to differentiate between one type and the other, it can soon get tricky. Some of the associations that could seem one type at the beginning can turn being the other type after studying them. For example, we could think that the size of an object and its sound are non redundant, but could turn to be that depending on the size of the object, its resonance changes systematically with size.

Then, we can say that crossmodal correspondences can be defined as the associations between two or more sensory features from different sensory modalities, being these features redundant or not. These correspondences occur between polarized features of the object. Generally all these associations are shared by a large number of people, and some are considered universal (we are going to talk about this in more detail later). To make this definition easier to understand or imagine we list a few examples of crossmodal correspondences studied before: between pitch and brightness (Marks, 1987a), where the higher the pitch is associated to brighter colors. Between pitch and shape/angularity (Marks, 1987a), where the higher the pitch is associated to more angular shapes. Between basic tastes and colors (Spence et al. 2010), and shapes (Deroy & Valentin, 2011; Spence and Gallace, 2011; Spence & Ngo, 2012; Spence & Deroy, 2014).

2.2.2. Types of Crossmodal Correspondences

This field, as we said before, is quite complex and has a lot of different versions and opinions from the people who study it. Depending on the person the conception and whole idea of the crossmodal correspondences is different, a big example is the classification of the type of crossmodal correspondences, that is due to the differences between researchers in considering where do crossmodal correspondences come from, if we learn them with the past of the years, or if there is a component that we have already before we are even born. After reading a lot and debating about it, we concluded that we agree with Charles Spence on his classification, where we have three different groups/types of crossmodal correspondences: first we have the Structural Correspondences, that exist due to the neural connections that are present at birth (Mondloch & Maurer, 2004; Wagner & Dobkins, 2009), secondly there are the Linguistic Correspondences, that are those ones based in the language. That means, when the words that people use to describe the stimuli are the same between the two senses. The easiest example is when we use the word high or low to describe a sound and this overlaps with the positions in the space when we talk about the location of an object. Finally we have the third group, where we find the Statistical Correspondences, in this group we have those correspondences that are learned during our lifes, through the co-occurrence (correlation) between two or more sensory events that are present in nature and in our surroundings.

As we can expect if we think about it, these last two groups may change depending on people's environments that means that they can vary depending on the culture of each individual, because they are based on experience and language. It could happen that some crossmodal correspondences that we can test and see common in some areas of the world or some cultures are different from the correspondences in the same field in other places. That is the reason why the cross-cultural study of crossmodal correspondences has started to be of interest of some researchers in the past years. That is why, based on this concept, we decided to work in the direction we've worked for this project.

It is important to distinguish crossmodal correspondences from synaesthesia. This is phenomenon in which, as in the crossmodal correspondences, one stimulus in one sense leads to an automatic, involuntary reaction or stimulus in another of our senses. People who have these experiences are called synesthetes, some of them experience just one, while others experience a big variety. A common example of Synesthesia is the associating colors to the days of the week. Another example a little bit more extreme is when the person experiences colors in vision while listening to music. Yet, whereas in cross-modal correspondences the association remains at a conceptual level, in synaesthesia the association involves a perceptual experience. That is, a color-grapheme synaesthete will "see" the colors associated with the letters, not simply think of the color (Grossenbacher & Lovelace, 2001).

Some researchers suggest crossmodal correspondences and synesthesia share more than what people would normally think. (Martino & Marks, 2001) said that both phenomena are based on the same neural mechanisms. Also another theory is that we all are in a line between normal and full synaesthetes (Martino & Marks, 2001). As we commented earlier on the memory for the crossmodal correspondences, this field is also really virgin, and need way more study to start having a better idea of the whole concept.

2.2.3. Are Crossmodal Correspondences Learned or Innate?

Nowadays the origin of the mappings between senses in crossmodal correspondences is unknown. Many studies have tryed to understand better how they work, but as other aspects in this field, it is still early to be sure of the origin of this phenomenon. We find two main theories to explain where do crossmodal correspondences come from: First, some people consider the idea that they are innate and therefore we "have" them when we are born. This theory, the physiognomic hypothesis, is based in some experiments and facts, for example the tendency of people to associate mood states with different color hues, as this phenomenon is not possible to be learned. Also, and even more important, is the fact that some experiments show how prelinguistic infants match certain type of stimulus with others in a coherent way, for example musical pitch with spatial position and visual brightness (Lewkowicz & Turkewitz 1980; Wagner et al 1981). In some other experiments, with other types of correspondences, though, is proved that this theory doesn't seem to work for all correspondences.

That leads us to the second theory, where experience is thought to be the base of the crossmodal correspondences and so, it is acquired rather than innate. The language plays a big role accroding to this theory, because some crossmodal correspondences, as we commented in the previous section, are based on it. So some of the correspondences are thought to not appear in humans till the point where language appears or at least till the person has a little bit of experience or is familiar in hearing linguistic sounds. A good example of the need of this theory is a paper by (Fernández-Prieto et al., 2015), where they show how the crossmodal correspondence between auditory pitch and object size seems to be present in 6 months old infants but not in 4 month olds. In this example we can see that in a small variation of age, there is a big change, and this is really possible to be due to experience.

Another good example is a study by (Ozturk et al., 2013), in this one when adults and infants were tested in a sound-shape mapping similar to "Takete" - "Maluma" experiment, both groups were reacting the same way, but when both were recieving similar words made with different combinations of only consonants or only vowels separately (therefore was a loss of symbolic phonemes and then there was a smaller

sound symbolism), the adults were still able to choose the proper correspondence, even if decreasing the number of "correctness", but instead the infants were not able to match the correct pairs. That is why it was concluded that we are able to see both theories in the same experiment. We can see the first theory in the fact that the infants have this crossmodal correspondences assimilated even if really young and precedent to the language learning and to any possible experience, but at the same time, adults, because of experience, are able to make this correlations increase and make them more "perfected".

The role of experience and innate aspects in cross-modal correspondences will be touched in our project, as we are going to try to add some more information to these two theories to try to understand better the origin of the crossmodal correspondences. We will discuss this further in the experiment chapter.

2.2.4. Cultural Diferences

Let's consider the wellknown crossmodal correspondence between pitch and size of an object, and go back to the definition of Statistical Correspondences, that are those which we learn by experiencing correlations in our natural environment. In our everyday life we perceive sounds from objects and learn how these are associated with the resonance properties so that the frequency at which a bigger object resonates tends to be lower than from that of smaller objects: Hence, no one is surprised that a kitten produces a high pitch "meow" and the roar of a grown up lion is, rather, low pitch..Hypothetically, would we live in a world with different physical laws, crossmodal correspondences would also adapt to other kinds of correlation. Yet, such hypothetical world cannot be empirically studied, and we have to point our studies in other directions. One major way to test this kind of idea is to study different cultures, that may have sounds the same way as we do, but that because of their surroundings, their everyday life, their language and basically of their experiences would be different and maybe would affect crossmodal correspondences.

Some researchers have focused in this idea to try to investigate and learn more about this phenomenon, they have studied cross-cultural effect on crossmodal correspondences of different sensory modalities and also the effects of culture in the

feelings perceived by the users with different types of technological design. A good example of the first is the study by (Wan et al., 2014) on "Cross-cultural differences in crossmodal correspondences between basic tastes and visual features", where they asked people from China, Malaysia, India, and the USA to match basic taste terms, to colors, shapes and textures. As a result they found important influence of the cultural background on matchings.

Another good example of the cross-cultural influences on crossmodal correspondences is a paper by (Bemner et al., 2013) on the "''Bouba'' and ''Kiki'' effect in Namibia. In this case the authors tested participants from a remote village in Namibia, with little exposure to western culture, on the "Bouba" and "Kiki" effect (reported before, as Takete-Maluma, (Kohler, 1929) and other less evident cross-modal correspondences. They found that these words are matched similarly as in the conventional way, but the results between other tests in relation with shape-flavour correspondences are answered the other way around (see Figure 2). We can see how in the carbonated water test, the participants related the sparkling water more with the rounded shape, while in western cultures the results are the opposite (Piqueras-Fiszman & Spence, 2012), also the test of bitterness with chocolate resulted the opposite way from western conventional results, where the more bitter the more related to angular shapes (Ngo et al., 2011).

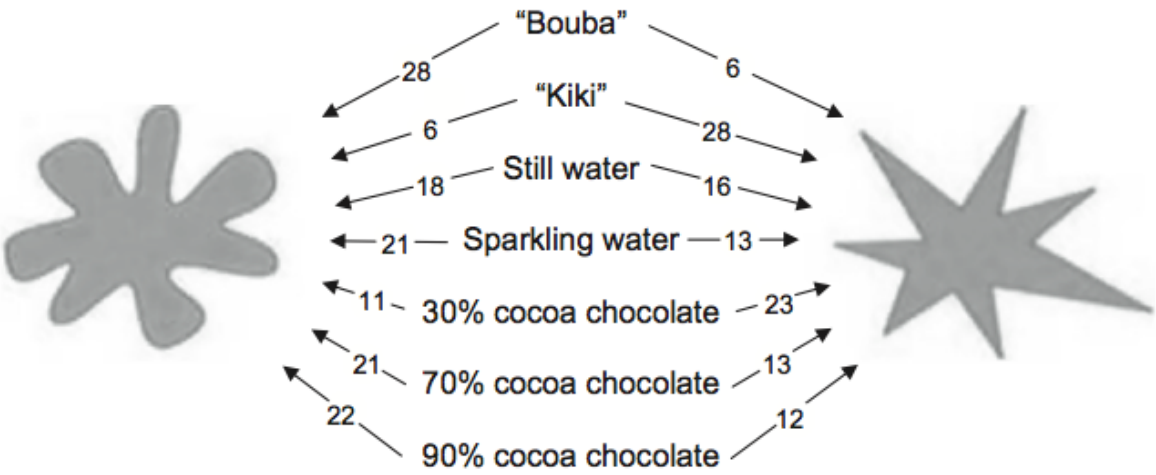

Figure 2. "Bouba" – "Kiki" Image From the Paper (Bemner et al, 2013)

As mentioned some lines above, we find other studies in direct relation with our project, and with the influence of the cultural background on the sensations that a stimulus generates on people. We have reviewed some work that focus on the connotation that colors have in different countries and therefore cultures. For example, while red was associated with the meaning of happiness in China, in the US was associated with danger. At the same time blue, was associated with a peaceful and relaxing color in the average of countries studied (Madden et al. 2000).

The idea showed in the previous lines is really important because it focuses on the knowledge of cross-cultural influences toward making better design decisions on websites, applications or general design, to make it be culturally congruent. Indeed, according to an ISO quality standard, localized design is a critical success factor for effectiveness and efficiency in task completion.

As companies become more globalised, the need for culturally sensitive design increases. To illustrate this idea, as of April 30, 2015 Coca-Cola operates a total of 112 different websites worldwide. Second, we have found a paper on "Cultural differences in the use of mobile devices" by (Leiber and Spanner-Ulmer, no date) that reports a study looking at the different ways of using mobile devices between China, Germany and the USA. The survey is based on the quantity of apps that the users run and the quantity of activities they do on the phone, to inform how they can make better localized designs regarding menu depths and the acceptation of new technologies.

Like we argue in this report, all these studies agree that if we have clear information of the variation of crossmodal correspondences in the different cultures in the world, the design decisions can include these localized correspondences, resulting in better, more effective and attractive technologies to the final users.

2.2.5. Multisensory Integration and Crossmodal Correspondences

Many studies show how information in one sensory modality (i.e., sound) can enhance perception in another sensory modality (i.e, vision). For example, it is easier to determine which temporal order of two successive flashes if one sound appears before the first flash and another sound after the second flash. This effect of enhancement has been attributed to the temporal information given by the sound about the visual information (Spence & Squire, 2003). Although this enhancement occurs at a sensory level (no meaning is involved), sometimes crossmodal correspondences that are mediated by meaningful relations may also produce multisensory enhancement. For example, the order of two quickly flashed squares of different sizescan be resolved more precisely if the surrounding sounds are in the right pitch-size congruency relation. The same situation could be done with other possible crossmodal correspondences, like lightness and pitch. These types of experiments illustrate one of our main interests and focuses on this project, the fact that crossmodal congruency can create distortion in the way we have perceptual experiences.

3. CULTURAL FRAMEWORK

In this section we are going to describe the situation in South Africa in terms of cultural diversity, focusing mainly in the language aspect.

South Africa is named the rainbow nation because of the large cultural diversity that exists in this country. Many are the cultures that over the years have been settling in this region and have made it one of the most heterogeneous countries in the world. South Africa has 11 official languages and some unofficial ones. The most commonly spoken language in official terms and public life is English, but is only the fifth most spoken in the country.

3.1. Language Distribution

According to the census of 2011, Zulu is the mother tongue of 22.7% of south africans, Xhosa follows with 16%, Afrikaans with 13.5%, English 9.5%, Setswana 8% and Sesotho 7.6%. The rest of the languages share the 5% left of distribution. We show a list from the census below to have a better idea of the numbers:

SOUTH AFRICAN LANGUAGES 2011		
Language	**Number of speakers***	**% of total**
Afrikaans	6 855 082	13.5%
English	4 892 623	9.6%
isiNdebele	1 090 223	2.1%
isiXhosa	8 154 258	16%
isiZulu	11 587 374	22.7%
Sepedi	4 618 576	9.1%
Sesotho	3 849 563	7.6%
Setswana	4 067 248	8%
Sign language	234 655	0.5%
SiSwati	1 297 046	2.5%
Tshivenda	1 209 388	2.4%
Xitsonga	2 277 148	4.5%
Other	828 258	1.6%
TOTAL	50 961 443**	100%

Figure 3. List From the Census of 2011

Most south africans are able to speak more than one of these languages, depending on their geographic area, they tend to speak more than one of the ones spoken where they live. Also, some of this languages have a lot of similarities, mainly because they have the same common ancestry. For example Zulu, Xhosa, Swati, and Ndebele are referred to as the Nguni languages.

To represent the language diversity we show the following image figure 4, where we can see painted in the south african territory map, the distribution of the different area dominant languages:

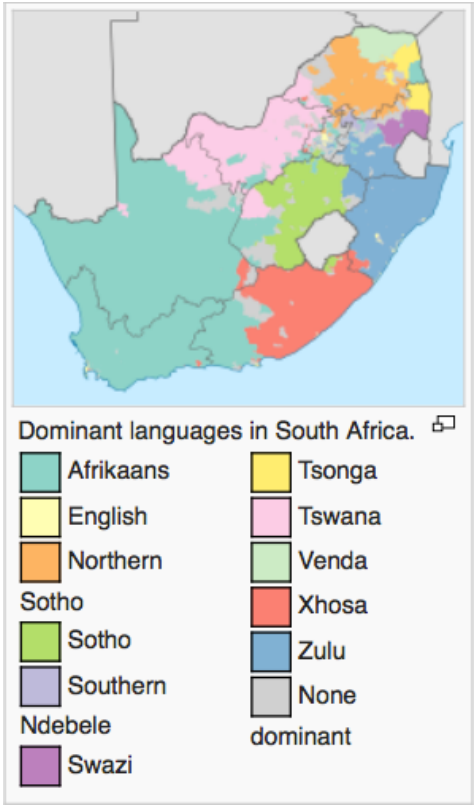

Figure 4. Dominant Language Distribution in South Africa

3.2. Western Cape, South Africa

This province is one of the nine provinces in South Africa, with it's capital in Cape Town. Situated in the south-west of the country is the fourth largest in terms of area and population, with an area of 129,449 square kilometres and 5.8 million inhabitants.

Figure 5. Western Cape, South Africa

The languages most spoken depend on the region of the country you are in. In this case in the Western Cape the situation is the following: Afrikaans (49.7%), isiXhosa (24.7%), English (20.3%). These three languages coexist in same áreas, influencing eachother in vocabulary aspects. Many of the inhabitants of the área speak more tan one of these languages. Also is important to know that the nature of the three languages is very different in terms of procedence and characteristics of them.

This heterogenic and complex situation as we mentioned before is really interesting for our project, because in a specific location of the Western Cape, we can find people really close from each other having a really different culture and language. (For more information about these three languages, their procedence, linaje etc, check the Appendix at the end of this document).

3.2.1. Mitchells Plain

Later on the experiment chapter, we will describe precisely to whom specifically we passed our experiments, to get the information needed to make our study, the type of people, gender, language etc, but firstly, still in this section, we are going to give a quick idea of the specific "town" where we decided to work in our project.

Figure 6. Mitchells Plain, Western Cape, South Africa

Mitchells Plain is a mostly coloured township in the metropolitan area of Cape Town, about 32 Km away from the city. Is one of the largest in Western Cape and South Africa, with 43.76 km^2 and an estimated population between 290.000 and 305.000 inhabitants. Built in the 1970's for the forced removal of colored people from the city, due to the apartheid, it was conceived as a coloured township, and even though nowadays is still the majority, there are other cultural groups and therefore many different background of the population of the area.

Group	Male	Percentage	Female	Percentage	Total	Percentage
Black African	12,692	4.48	14,021	4.95	22,723	7.32
Coloured	122,360	43.17	131,485	46.43	281,829	90.77
Indian/Asian	902	0.32	861	0.30	1,926	0.62
White	274	0.09	307	0.10	581	0.19
Total	151,033	48.64	159,453	51.36	310,485	100

Figure 7. Race Distribution in Mitchells Plain

The distribution of the “first language” spoken in the township is the following: English 47.4%, Afrikaans 46.9%, isiXhosa 3.3% and 2.4% of other languages.

Some neighbourhoods of Mitchells Plain used to be in way better shape that they are today, with the past of the years the situation of the township has been decreasing, and has made a lot of the areas become ghettos, where gangsterism, drug abuse and informal settlements have increased in a drastic way.

Most of these situations described, even if not comfortable nor idyllic at all are quite interesting and essential for our project. The multicultural aspect, plus the bad economic situation of the population that implies, in general, a decrease on the percentage of people owning technological devices like smartphones or computers and that means the less experience or familiarity with this technologies, plus the distance from the city of Cape Town, and therefore the lack of contact with the most occidental city of the region, make this specific location be of a big interest for us and our project, to try to demonstrate the ideas that we have.

After the general idea descrived during this section, about the situation in South Africa, Western Cape and Mitchells Plain, both in economic and cultural aspects, we can say that the area chosen to do our study and experiments is one of the best regions possible worldwide to focus on cultural differences in crossmodal correspondences and be able to study the possibilities of cultural adapted design.

4. EXPERIMENTS

As we discussed in the introduction, our main goal is to collect information on crossmodal correspondences to be able to have a better idea on how people from certain cultures or areas feel about the stimuli they receive. That is useful for us to be able in the future to generate a better design in technology, on how to create something more pleasant and attractive for the users, and that could be done by generating a type of collaborative design[1], and that is the idea we followed to generate our experiments.

We decided to experiment the ideas we had, on some individuals, to see how the different concepts described before were having effect on people, if the way it was expected or differently. That is why we created a serie of tests that we are going to describe in this chapter, with the ones we tried to get the maximum information we could in this field and in the direction to find answers to the questions descrived previously, like if the crossmodal correspondences are innate or learned, if there are differences in percieving crossmodal correspondences or not, etc.

As mentioned in previous sections, our experiment is intended to be tested with children of different ages and two different cultures (we will be more precise about the participans in the Methods section of this forth chapter), this is because we think that this caracteristics, are the best posible to find out about crosscultural variations in perceiving crossmodal correspondences, and also and mainly, because the fact of being children, with a limited experience, helps us understanding if experience plays a big role on crossmodal correspondences and at the same time is usefull for understanding how to make an intuitive design, with participants that don't have the experience field and arn their decitions are not influenced by previous experiences.

We generated a simple and easy test series, inspired on previous studies, but adapting them accordingly to our interests and the characteristics of our young participants. With these tests, based on crossmodal correspondences, we wanted to see how the cultural and age differences affected on the perception of simple sensory correspondences, to apply these results on possible future design.

[1] Also know as participatory design, is an approach to design, involving actively the stakeholders in the design, to make sure this one meets their needs and is usable.

Basically we wanted to test the universality of the crossmodal correspondences chosen by testing these in a novel culture or population (previously untested), and compare them with the results obtained of the same tests done in a western group of participants. We were quite sure that some of the tests would have similar results between both cultural groups as with similar tests taken in other areas of the world, but at the same time we expected that other tests would have variations in the results between the different cultural groups. At the same time, we were expecting differences in the results, between the different age groups tested, due to the possible role of experience, for at least, some crossmodal correspondences (as discussed in the sections above).

Also, due to the complexity of the situation, the exact place where we decided to make the test, the testing conditions in terms of cultural environment, but mainly the age aspect, made us spend some time consideration how to make these tests accessible and comprehensible for the participants chosen.

4.1. Chosen Types of Crossmodal Correspondences

We choose a few crossmodal correspondences to base our study on, that we consider the most relevant, interesting and that the results could be interpreted clearer to apply on our goals. We could have chosen many other types of correspondences but mostly we choose them on the field of visual correspondences in relation with sound, because they are far more studied than the rest, and because it is more likely that auduio-visual correspondences will be applied in technological design (than, for example, taste or smell based correspondences). Even though we consider that the more recent studies in the correspondences between taste/olfaction and vision, for example, are something really interesting for the future of design of user experience, due to the many possibilities that it gives. The crossmodal correspondences between sound and vision, are the ones more studied historically, and that is why we could have more information about past tests and past conclusions if we were working in these same field that if we were getting into newer paths.

We tried to make a choosing that would help to see in the results the different points discussed before in the “Theoretical Framework”, for example the aspect of if the

crossmodal correspondences are innate or learned. That is why we included two correspondenes that we consider totally arbitrary (the “Takete”-“Maluma” effect, and the visual lightness and pitch), and two others that we consider that could be learned from experience (visual spatial position and auditory pitch, and visual size and auditory pitch). Indeed, we chose the different correspondences thinking of the different types described a few sections before; Structural (visual lightness and auditive pitch), Linguistic (visual spatial position and auditory pitch) and Statistical Correspondences (visual size and auditory pitch), in order to give a better idea of differences between correspondences and to see how this in relation with the learned of innate discussion described before affects our participants.

Finally we added two further correspondences of a type that we decided to call “technological crossmodal correspondences”.(scrolling up and down on a computer screen or tactile device and zooming in/out in the same surface). Basically, we consider these type of correspondences those that we can say that they are directly associated with technology. For us, these correspondences are important because they are specifically relevant to the finality of this research that is to find out ways of making a better design for technology with a big interest in the interactive aspect of ti, and how could this be learned better than by testing the potential users of this technology with this correspondences?

The correspondences chosen are the following:

- **Pitch-Lightness**: There is a big background of research already done, in different groups around the world, but mostly the participants were occidentals (european or american), and they didn’t test the universality of the correspondence, at least within the cultural group we did. If we go back the classification we discussed of crossmodal correspondences, we would consider this one to be Structural, because there is no physical rule, that we could get by experience that would make our brain “memorize” that the higher pitch associates with lighter colors or, the opposite. And we consider it to be an association intrinsic in our brain pre birth, because there is no other possibility that fits in our classification, neither a way that we can imagine of learning this correspondence.

If we go back a few sections and we remember the classification we did of crossmodal correspondences, we would consider this one to be Structural, because there is no physical rule, that we could get by experience that would make our brain "memorize" that the higher pitch the lighter or the opposite. And we consider it to be an association intrinsec in our brain pre birth.

Previous studies are: the one done by (Bond & Stevens, 1969) at Harvard, where they showed than adults and 5 years old children matched brighter surfaces with louder sounds (J. C. Stevens & Marks, 1965; Root & Ross, 1965). Louder sounds with higher contrast images (Wicker, 1968). Hue and pitch in children, where high-pitched tones were more likely to be matched with yellow color rather than with blue (Simpson et al. 1956). We could list many more examples of similar studies of crossmodal correspondences close to the one between pitch and lightness, many studies between auditive pitch and a visual color/light variation but the ones we are really interested are the pure related to our test, where we find the relation between brighter surfaces and high pitched tones (Marks, 1974; WIcker, 1968) and finally between higher pitch and lighter surfaces (Marks, 1987; Melara, 1989).

These last past studies all show the correspondence between pitch stimuli and lightness surfaces. And they all conclude that the participants answer accordingly when higher pitch is associated with lighter surfaces. By having this conclusion, clear and common in all the major studies in the field, we wanted to test if something as obvious where it was tested, was going to be the same in the age and cultural groups we decided to test.

We include in this section a little abstract in chart form to have a better idea of the main studies done in this field in the past, most of the ones we used as an inspiration or as resource of information. As it is possible to see, all of the studies listed below, are between visual and auditory stimuli, but they are not all the same type of correspondence as between pitch and lightness. We use this list to inform of the variety of correspondences and studies on this field. All these results collected in these charts

are obtained using the speeded classification task[2], other results gotten by other types of tests are not shown in this image:

Auditory Dimension	Visual Dimension	Crossmodal Correspondence	High-Pitch/Loud Sound Corresponds to:	Studies
Pitch	Elevation	Yes	High elevation	Ben-Artzi and Marks (1995); Bernstein and Edelstein (1971); Evans and Treisman (2010); Melara and O'Brien (1987); Patching and Quinlan (2002)
	Brightness	Yes	Brighter stimulus	Marks (1987a)
	Lightness	Yes	Lighter stimulus	Marks (1987a); Martino and Marks (1999); Melara (1989a)
	Shape/angularity	Yes	More angular shape	Marks (1987a)
	Size	Yes	Smaller object	Evans and Treisman (2010); Gallace and Spence (2006)
	Spatial frequency	Yes	High spatial frequency	Evans and Treisman (2010)
	Direction of movement	Yes	Upward movement	Clark and Brownell (1976)
	Contrast	No	N/A	Evans and Treisman (2010)
	Hue	No	N/A	Bernstein, Eason, and Schurman (1971)
Loudness	Brightness	Yes	Brighter stimulus	Marks (1987a)
	Lightness	No	N/A	Marks (1987a)

Figure 8. List of Crossmodal Correspondences Experiments Vision – Audition (Spence, 2011)

This crossmodal correspondence, being part of the group named Structural correspodences, that meaning that is not a correspondence learned, therefore that experience plays a big role, is really interesting to be tested, to try to prove the aspect of the experience. That is why experiencing with children, makes this correspondence a good choice, to try to understand how the experience in a group of people, that are really young affects or not on the connections they make between the different senses, being in this case audition and vision.

- **Pitch-Size**: This crossmodal correspondence is also one of the most studied. We find different previous studies in relation: Edward Sapir saw the existence of a correspondence between the speech sounds /a/ (lower pitch) and /i/ (higher pitch) and the size of an object, where nonsense words like "Mal" and "Mil" were associated

[2] Type of task where the participants have to make a decition between two or more posible options in a really short period of time. This type of task makes the experimenter recieve the impulsive response from the participant instead of a reasoned answer.

accordingly with this idea, “Mal” with bigger objects and “Mil” with smaller ones (Sapir, 1929). Another example is that 3 year olds matched high pitch tones with small bouncing balls and low pitched tones with big ones (Mondloch & Maurer, 2004). And also we find previous studies in the same exact matter: 9 year olds matched high pitched tones with small objects and low pitched tones with big one (Marks et al., 1987). And as shown on Figure 8 X, tested with speed classification we find the relation between high pitch and small objects (Evans and Treisman, 2010; Gallace and Spence, 2006). An example of test done with the speeded classification task, to test the pitch - size correspondence is shown in the following Figure 9:

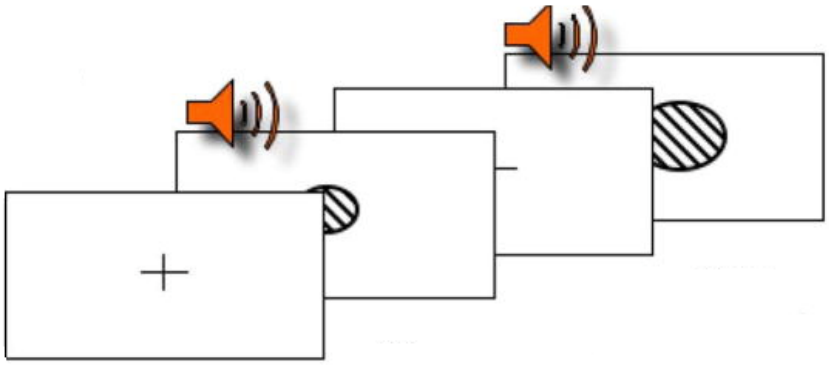

Figure 9. Example of Speeeded Classification Pitch – Size Correspondence (Evans, 2010)

The pitch-size correspondence is classified as a Statistical correspondence, because, as described in previous sections, this is a correspondence that is thought to be learned through experience. In nature, in the environment surrounding us, we find that the frequency of resonance of bigger objects is lower than the one from smaller objects that makes that in our brain we have learned that when we hear a sound with a high frequency we are going to associate that sound with the smallest of the objects. This kind of correspondence has been demonstrated, amongst others, by (Gallace and Spence, 2006) using a speeded classification task where the participants responded to disk of different sizes faster when the pitch of a concurrent sound was congruent (i.e., big disk low pitch), versus when it was incongruent.

In this case, we thought, that even if being a correspondence ruled by experience, and even if the cultural situation was really different, we didn't think that the natural environment would change that drastically between the Western Cape in South Africa and the regions where this kind of test was done before in the United States of America or Europe.Even though we tested this correspondence to see the results of a different culture, and to see the variation between ages, to try to get some information that could surprise us or give us some directions in our conclusions.

Even though we tested this correspondence to see the results of a different culture, and to see the variation between ages, to try to get some information that could surprise us or give us some directions in our conclusions.

This example shows how this crossmodal correspondence could be used in an interactive design. Making the interaction easier and more intuitive for the user, using this relation between senses. That is why we considered this correspondence as one of the important ones we wanted to test.

- **Pitch-Vertical Position (elevation):** This correspondence refers to the association between high (low) pitch sounds and visual events at high (low) spatial locations. Some of the most relevant studies done before are the following ones: (Bernstein & Edelstein, 1971) were the first ones to demonstrate in a study that people respond slower to a visual stimulus if it is incongruent in terms of spatial elevation with the pitch of the sound that is played at the same time as the visual stimulus appears. Also (Cabrera & Morimoto 2007; Roffler and Butler 1968; Rusconi et al. 2006) demonstrated the same idea using different methods. This correspondence also has been tested by using sweeps (i.e., high to low sound) to guide the attention of the participants (toward low position, in the example), with young infants with a mean age of 11.4 months old, that were matching up or down arrows with pitch sweeps up or down respectively (Wagner et al., 1981). A schematic example of the tests done by speed classification is represented in the following Figure 10.

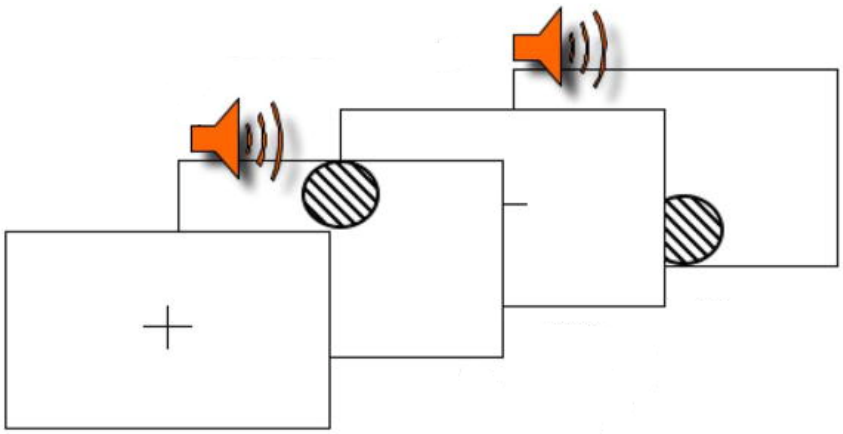

Figure 10. Example of Speeded Classification Pitch – Position Correspondences (Evans, 2010)

This crossmodal correspondence is classified in the Linguistic group, because is thought that the association comes from the language, as we use the same adjective to describe tonal height than as elevation in space, and that makes our experience tell us that a high pitch should be in a high position.

This last paragraph makes us understand also, the interest of this correspondence to be studied in a different culture and in different groups of humans, because as we said, this correspondence is dependent on the language, so if the culture and therefore the language are different, the correspondence could change from culture to culture, making it be equal or totally different from one to another. This is why the results on this test could mean a lot.

Even if most cultures match the high pitch tone with high elevation, it was demonstrated that certain Native American populations were not having this response to the stimulus (R. Walker, 1987). This shows clearly how our intention is interesting and could have a surprising response. Also, knowing that there are different possible answers makes it more interesting for a future design that could be adapted to each cultural group. This idea, could be really helpful, as the ones before, to guide the attention of the users of an interactive or another design, to make the use of the device easier more intuitive, faster and more harmonic.

- Maluma – Takete: this is one of the most famous correspondences, the biggest example of sound symbolism, that has been tested by many researchers, in different circumstances, regions of the world, groups of population and with many variations. Is one of the clearest examples to understand what a crossmodal correspondence is. How the words "Takete" or "Maluma" can be associated to an arbitrary shape, rounded or angular shows how this process works.

The first to test this concept was Köhler (1929). There are many theories to describe why this phenomenon happens, some say that the sound that an angular object does when hit or thrown to the ground is more similar to one of the words than to the other, other theories are related to the combination of letters and consonants in the words that as studied before, some letters provoque a different association than others.

This last idea is based also in many studies made, understanding the cross cultural effect of this experiment. These cross-cultural studies showed little differences between groups of population, with variations in the sound symbolism effect (Davis, 1961; Diffloth, 1994). During the years, this particular study has evolved, and has been changing and adapted to different cultures, ideas or goals. For example, the words used to describe the shapes started as "Baluma" and "Takete", as tested by Köhler (1929) in Spanish population in the Canary Islands. After a few years, in his 1947 book version, Köhler changed the word "Baluma" for "Maluma", answering like this to the concerns that the first word was sounding too much like balloon and that could guide the test participants reactions.

Afterwards, (Ramachandran & Hubbard, 2001, 2003) replicated Köhler tests by changing slightly the words, using in this case "Bouba" and "Kiki" and got similar results. There are many other examples in relation with this test, for example when english children between 11 and 14 years old, and Swahili-speaking children between 8 and 14 years old, from central Africa where asked the same question with the words "Takete" and "Uloomo" both groups answered similarly to the question.

Also, this type of test is really interesting for other sensory modalities, that could result as something really interesting for future studies or future applications, in the fields of marketing for example. The paper "''Bouba'' and ''Kiki'' in Namibia? A remote

culture make similar shape–sound matches, but different shape–taste matches to Westerners" by (Charles Spence et al, 2011), is a really good example. They tested the effect in a remote culture in Namibia and extended it to associateions with flavours. So for example, they tested which shape the participants were associating to sparkling or still water, and they got surprising responses, different from the ones usual in tests to occidental cultures.

The reason we choose this correspondence is the fact that is widely studied, and so interesting, and we wanted to make sure, even if it was already quite obvious, that in the specific cultural environments we were making the tests it was having the same results. Also we wanted to make this experiment because we find the possibilities of it really interesting for future designs and future applications, and as shown in the past paragraph, the results of the test variations could be used in many different fields. The one most obvious is the marketing one, where words can be used to make a brand look sharper or bolder, while others could make a brand or a slogan be softer. Also, some shapes could remind some flavours more than others, or some characteristics of tastes, smells etc.

- Scroll Up-Down: as we commented in the previous section, we thought, that apart from the crossmodal correspondences that we found in papers and in many studies, we would like to test something else, more pointed in the direction of our project. Something more specifically related to the future design better design of an interactive, or a web design or any technological design adapted to a culture. Therefore we thought that going directly to some interactive patterns we use to interact with our smart devices, computers or other technology would be the right thing to find some interesting correspondences that would be good to see how the cultural difference affects on them.

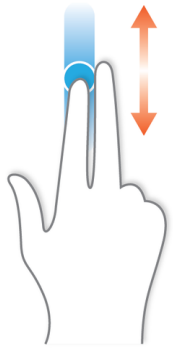

Figure 11. Scroll Up – Down Gesture [1]

That is why we found the most relevant one in our opinion. Being users of Mac, for a few years, we have seen how with the recent update of the operating system (from OS X 10.9.3 onward) the effect of scrolling up/down, has changed. Before the OS X version, to go down on a page the user needed to move the fingers from top to the bottom of the track pad, but nowadays is the opposite. For us the past concept was really clear and natural to be done this way, but with the change it turned to be in our opinion quite anaesthetic, till the point that makes the interaction somehow unattractive. At the same time people that started using Mac later on when this change was applied already, most of the people asked were argumenting that it was really natural for them and a good way of interacting.

Due to the situation described before, we wanted to understand how this controversy could change between cultures, to see if in a culture it would be more natural doing it one way than another one. So we thought of testing this correspondence with our participants to understand their preferences, in the way of a collaborative design, where depending on the answers or opinions of the potential users, you adapt the design to their preferences.

Also the fact that the group we decided to test was in a cultural and surrounding situation quite particular was making this kind of experiment more attractive. The participants we chose had a null or really small experience with smart devices or computers, due to the poverty of the area where they live, the situation of their families

and the small influence by the closest big city, Cape Town, so that made that if we were getting consistent results, that would be able to be understood like the most natural way for this group of population to do this gesture.

Zoom In-Out: We considered also the action of zooming in and zooming out as part of the technological set of correspondences that we have discussed above. We think that is an interesting action that the results could make a difference for a future perfected design, adapted to the background of the cultural group tested. In this case we also have an action that we consider more or less arbitrary, and that interacts between vision and touch trying to mimic the idea of making a part of a picture, page or similar bigger or smaller.

Even if in this case we think that the association is clearer and there is not as much possible problem as with the previous one we thought that we could have the concept really inside of us, but for other cultures could be totally different by experience or even more if the potential users don't have any experience in the field. That can make the results more interesting, because could show how our idea can be understood differently by other people.

Therefore we think that this test, as the past one is really interesting because could inform us of something that could be done better or adapted to a type of people or age group, that could help everyone on that group have a better performance with the interactive device, and even more if the group of people is starting from zero using technological devices with this type of interaction, Could help them learn easier the ways of using it, making it more intuitive and confortable.

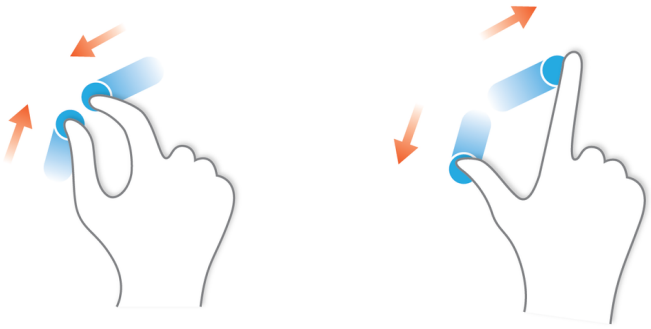

Figure 12. Zoom In – Out Gesture [1]

4.2. Methods

4.2.1. Participants

For our experiment we had two main groups of participants, the first group was integrated by a group of kids from the Xhosa tribe, all part of the school "Keep The Dream" in Mitchells Plain, Western Cape, South Africa. A township where the condition of most of the families and therefore of the kids is quite difficult and the poverty is present in most of them. This situation plus the fact that the area is away and quite isolated from the biggest and most occidental city around affects in a way that the inhabitants of thisi township are less familiar with technology and also and really importantly less influenced by the occidental culture. None of the participants of this group had ever done a test of research or anything similar to what we did.

30 kids from the Xhosa tribe (12 male, 18 female) were selected from a total group of 73 that were forming the school, with the help and knowledge of the teachers and what the experimenter was perceiving while interacting in normal school activities with them, and afterwards they were tested. We divided them in two different groups depending on two main factors; the first one was the knowledge of English and the second one the age factor. These two aspects were most of the time in strictly clear connection.

From the total number of students, we selected 15 for the first group. These ones were between 6 and 7 years old and they were having enough knowledge of english to be

able to follow instructions given to them and to answer back in english, but their original language and the one used to communicate between them was Xhosa. The second group with also 15 kids selected was way different. In this case the participants were younger, between 3 and 4 years old and in clear relation with that, their knowledge of English was null. From the youngest kids in the school, we chose those ones that were having less knowledge of English, and they were mainly only able to speak and understand Xhosa.

From the first age group, with older kids, two of the participants were excluded of the interpretation of the results, because they were having too much experience with smart devices, and we thought it would not give us the relevant information we were looking for. From the second age group, a participant was also excluded, but because of other reasons. This participant was not able to pass the first simple test of association of images and sounds to verify that the participants would be able to follow the rest and more complex tests.

The second main group, was integrated by a group of kids from Escola Tres Pins of Barcelona, Catalonia, therefore the kids were considered being part of the western culture. Having then a total different background and situation from the first group. Living in a big and evolved city, in continuous touch with technology and a western type of life.

30 kids from this school were also chosen (14 male, 16 female), selected from the total number of students that were forming both classes, “1r de primaria” and “P4”, with the help of the teachers and the impressions of the experimenter while interacting with them during some activities while in the class. All the kids from this second group were speaking catalan as a mother tongue and they were tested in this language. As in the previous group, we divided the students in two age subgroups, divided in 15 kids between 6 and 7 years old and 15 between 3 and 4. Differently from the african participants, the difference between both age groups were not as evident in terms of language and understanding.

In this group, no kid was discarded from participating on the test nor their results were not used to be evaluated.

Due to the condition of these two groups, we had to prepare and execute our tests really well, thinking of all the characteristics to be perfectly adapted to them, for them to be able to follow the instructions and give us some interesting information in their answers with the one we could work. We will talk more about this matter in two following two sections.

4.2.2. Apparatus and Materials

The participants were presented in front of the experiment differently depending on which of the previously described groups they were part of. As mentioned before due to the early ages of the participants, we had to prepare well all the parts of the test if we wanted the kids to be able to react to the experiments properly. That is why we had to think of all the different aspects involved on the experiment and their surrounding; the type of test, the possible distractions, the familiarity with the surroundings, the understanding of the instructions given by the experimenter and others.

The participants from the first group that were able to communicate in english, that means the older subgroup, were placed in a silent room, familiar for them, away from the other kids from the school, with the experimenter and a teacher close to them just to give them a sensation of security and make them feel safe, the teacher was not interacting with the kids unless strictly needed. Differently, the younger subgroup of the south africans, not able to understand or speak english were situated in the same room as mentioned previously, but in this case, the teacher that was with them, was closer and was interacting constantly, because was working as a translator.

The kids from the Catalan group were also placed in a separate room, one by one, away from the rest of the students, to be tested. They were also having a teacher in the same testing room to give confidence to them and be able to be more relaxed while working on the experiment. The difference in this group, differently from the previous one, is that both subgroups divided by age were tested the same way, because they were both speaking good Catalan and were able to communicate perfectly with the experimenter without needing anyone else.

A serie of images and sounds were presented simultaneously to the participants, before each test, the experimenter was asking the participant a question to be answered after each couple. The question was always to associate one of the two sounds played with one of the two drawings presented in the same image. The images were presented in the 15 inches screen of a computer, covering the whole screen. The keyboard was covered with a white paper to avoid any type of distraction, for the participant to be focused on the image appearing only. The sound was played from the same computer in a proper level of sound for them to be able to hear it properly and hear the differences between the different sounds presented.

As we will comment in the next section, each image was associated with a pair of sounds, and that was presented to each participant. Our experiments were adapted to the age of the kids, to make them see them more appealing and similar to a game, to make them want to do it better. Normally similar tests of association of sound and image are made with shapes or other abstract objects, but in this case we wanted the kids to feel more comfortable with the experiment, so we used in most of the different parts of the experiment animal drawings. And therefore the sounds related to this drawings were animal sounds.

There was not an specific order on the questions, images and sounds that were presented except for the first one and last two groups, because the first test was supposed to indicate if the participant would be able to follow the next tests and the last two, because it was an extra abstraction step after the previous part of the experiment, that we thought it would just be more natural to do it later on time. The different tests of the middle part of the experiment, were presented aleatorily to the participants, on purpose to avoid any possible results that would come from just the order of the tests were presented. So this way we tested some participants with some images first and some other participants with other ones.

The first image, as we commented in the previous paragraph was the same for all the kids. It was presented in the whole screen and afterwards, with the image still showing, four sounds of animals were played, one by one, for the kids to associate each sound with each animal. The image was the following:

Figure 13. First Test

The four drawings contained in this first image were the ones used, to represent the next parts of the experiment. After this first test, if the participants were passing it properly, they were getting shown on screen, in the same way, more images in different orders, in relation with the rest of the tests.

In the previous section "Chosen Type of Crossmodal Correspondences" we explained as the title recalls, which correspondences we decided to test. That is the way we decided to group our tests, to try to get the information from the participants, for each correspondence different ways, to give more base to the answers they were giving, and to realize if there was consistency on their answers or it was just an aleatory game for each of them.

Therefore we had three main groups in this second part of the experiment, after the initial test, divided in three crossmodal correspondences type. To choose one we can start by the correspondence between pitch and visual size. In this case as in the others two, we plotted the same idea represented by the four animals shown in the previous image, as we said before, to give consistency to the answers of the participants, to make sure that the participants were not having a bigger affection to one animal than to another one and that was ruling their decisions.

At the same time, by each animal, we generated two images counterbalanced, so if we were presenting a plot with a small pig on the left and a big one on the right, we were also having another plot with a big pig on the left and a small one on the right. This was used also to give more consistency to the answers, and to make sure that the crossmodal correspondence that we were trying to test was not affected by the position of the images on the plot. We can see this idea in the next two figures, Figure 14 and Figure 15:

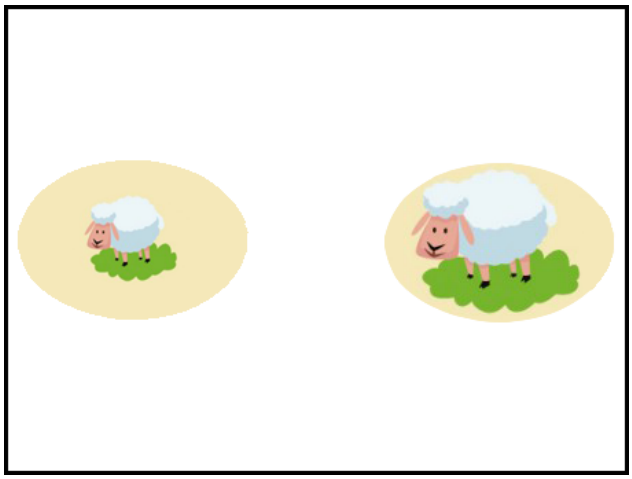

Figure 14

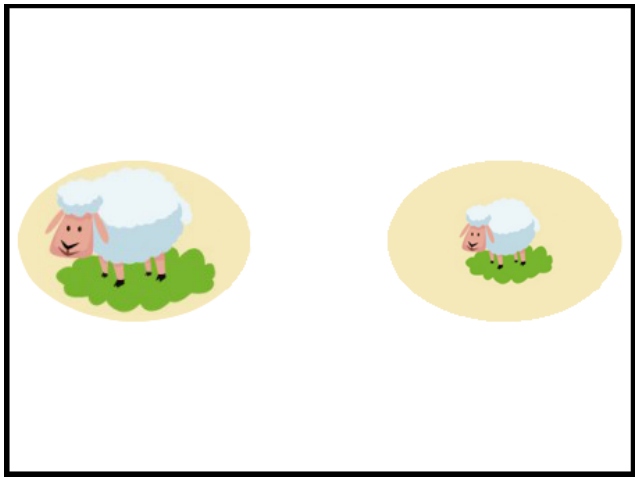

Figure 15

As said before, for testing the pitch-size crossmodal correspondence we were having four different possibilities of the test, each one with it's counterbalanced variation, they

were shown to the participants aleatorily, for one of the participants we were starting by the cat, and following by the dog, showing the four of them, while for another participant we were starting by the pig and keeping going.

As mentioned in previous sections, studies done before show how participants match higher pitched sounds with smaller objects and lower pitched sounds with bigger ones, with this design that we created we followed this idea to try to verify if this situation was also shown in our participants.

Using the same method explained previously for the first selective test, the participants were seeing shown on the whole screen one of the following images; Figure 16, 17, 18, 19, and once each image was appearing two sounds were played one after the other, while the participant was asked to match each sound with each of the two animal variations of the image. The two sounds played were, for each image the pertinent animal sound in two different pitch, one low pitched and one high pitched. The pitch different between both sounds was always enough for an average ear to discern between them but still be able to understand the origin of the sound, that means understanding that the cat sound was actually a meow. The different images for this test were the following (excluding the counterbalanced designs):

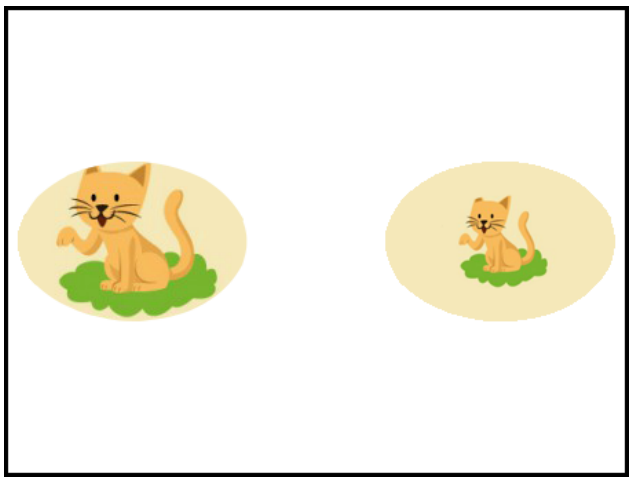

Image 16

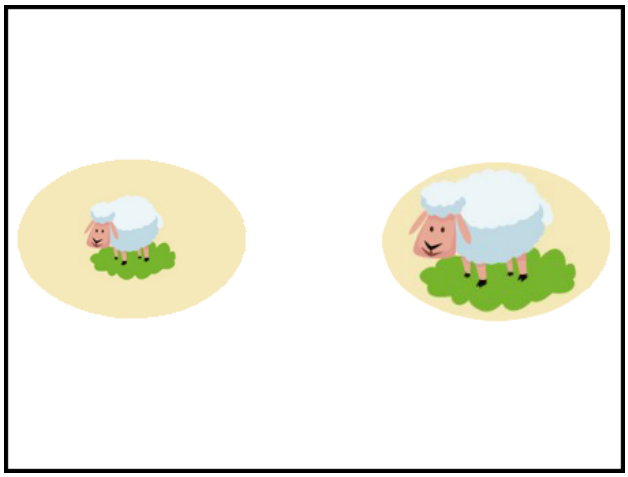

Figure 17

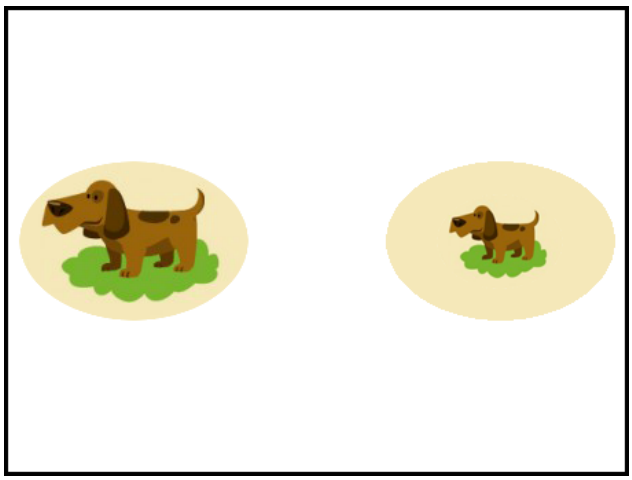

Figure 18

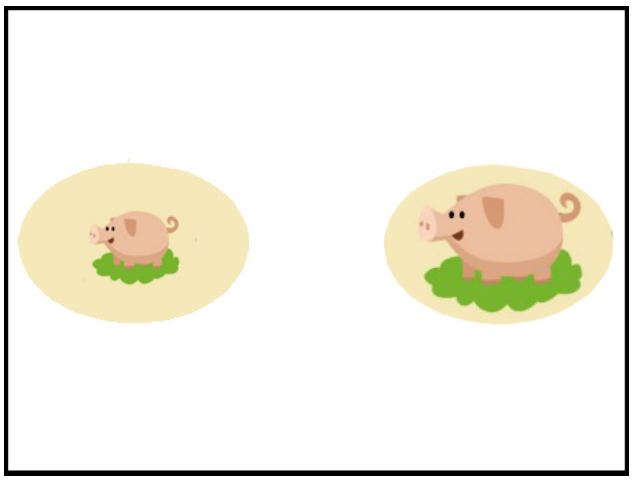

Figure 19

Another of the three crossmodal correspondences left in this middle part of the experiment is the relation between pitch and lightness. In this case as the previous one we have the four animals with their counter balanced variations.

As explained in previous sections, some studies show how participants match higher pitched tones with lighter images. We generated these images following that idea, to try to prove by ourselves that our participants from the specific culture and area they are from were having or not the same reaction.

In this case also, the participants were getting shown on full screen the images of this group in an aleatory way and once each image was shown, two sounds related to that image were played. Also the two sounds were the sound of the animal shown on the plot, we actually used the same exact sounds as the ones in the previous test, one animal sound high pitched and the other low pitched. The participants were also asked to match each sound with each of the two animals in each image. As we can see in these following images Figures 20, 21, 22 and 23, we used the same based design as the previous examples, but applying some variations to adapt them to the actual test.

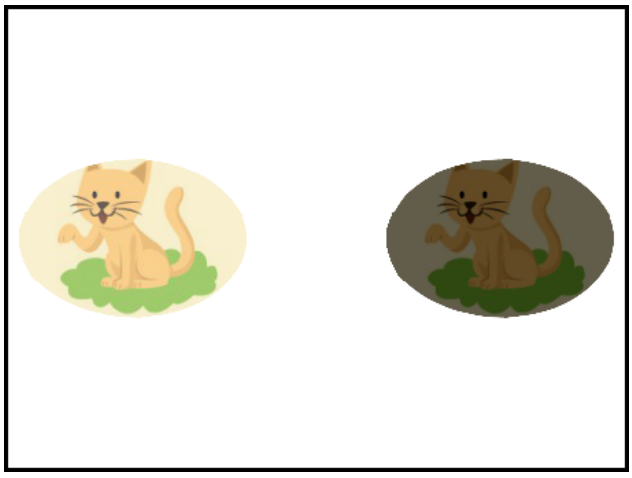

Figure 20

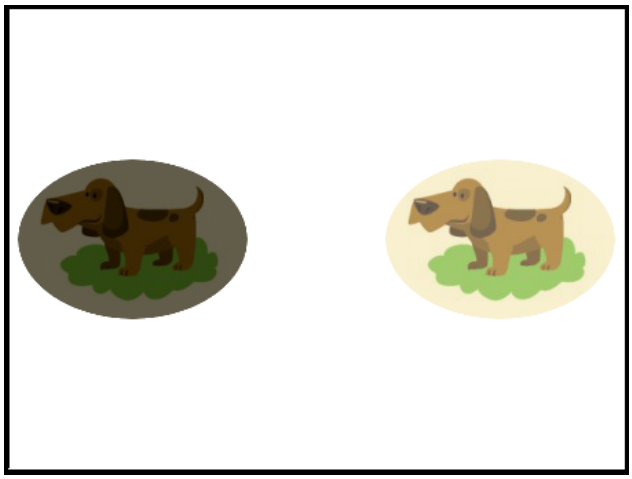

Figure 21

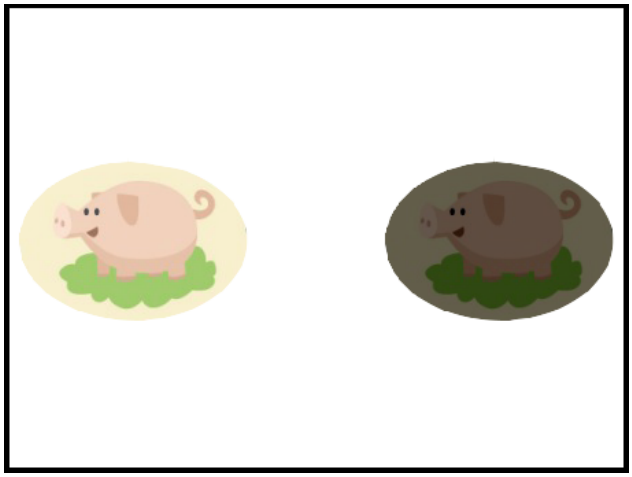

Figure 22

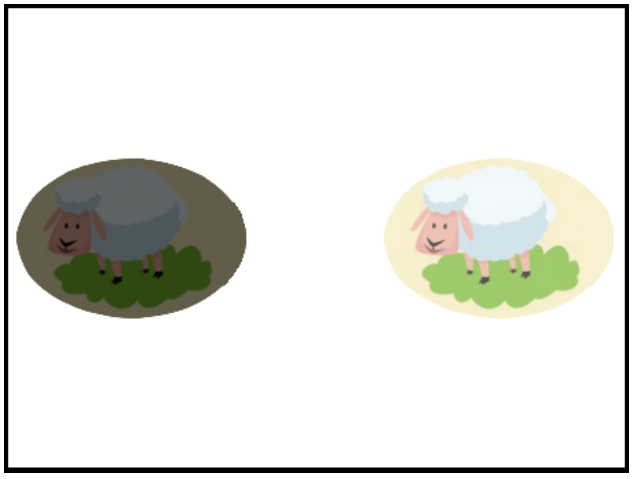

Figure 23

Lastly we have the last crossmodal correspondence of this middle section of the experiment, the relation between auditory pitch and spatial position. As for the past two tests, in this case we used also the four animals with their counterbalanced variations. We also used the same design base but with the pertinent variations to test this particular correspondence. We also included in this case a cross in the center of each image, to give to the participants a reference point, that we thought was needed for this specific test.

Some tests done before by other researchers show how pitch or sound sweeps can guide the attention into different directions, we explained this in previous sections. That is why we decided to generate these figures, 24, 25, 26,and 27, to understand better how this particular participants were reacting to this matter and this specific design.

As for the previous sections, the participants were getting each image in full screen size (also in an aleatory order), and while the image was showing two sounds were played and the kids were asked to match each sound with each animal variation of the image. In this case though the sounds were based in the same animal sounds, but they were transformed into two sweeps in each case. So for the dog, there was a sweep going from a lower pitch to a higher pitch and another one in the opposite order, and the same for the rest of the animals. The images used for the test of this correspondence are the following:

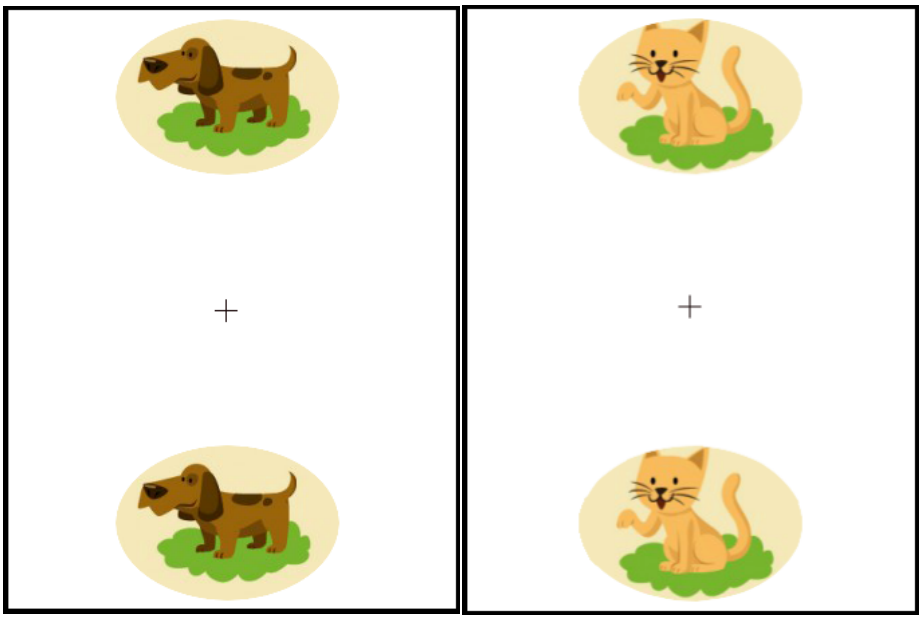

Figure 24-25

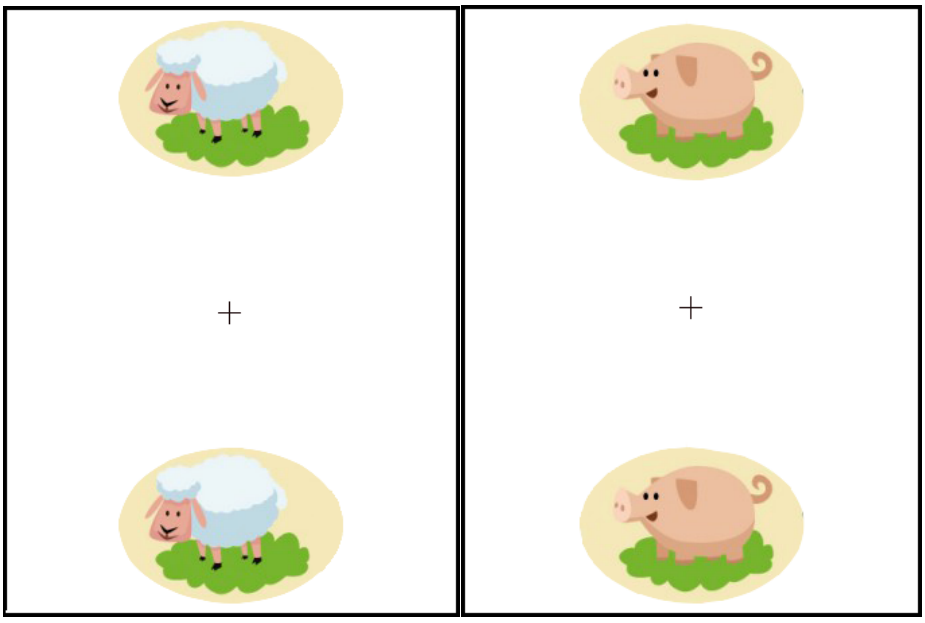

Figure 26-27

The third part of our experiment was consisting of going a little bit further on abstraction. Once the kids were more familiar with the procedure to follow to pass the different tests of our experiment, due to the more playful and attractive tests with the animals, we wanted to see if the results from these firsts tests would be similar with

similar tests but in this case with shapes instead of animals and pure tones instead of animal sounds.

The main idea was to do the same exact three tests plus an extra one, but in this case with geometric shapes, that is how most commonly researchers test this kind of experiments. We designed 3 images with the same concept as the previous tests but as said with other objects on them. For the pitch-size crossmodal correspondence test we made the following Figure 28. In this case the two sounds used to play while the image was showing were to variations (one high pitched and one low pitched) of a pure tone, also in this case the difference between the two variations was evident for an average ear, so the participants were not having any problem on differencing one from the other:

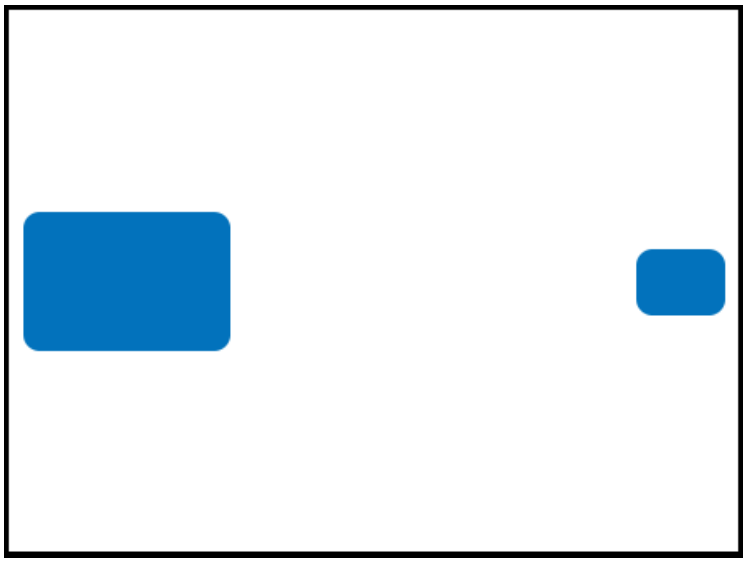

Figure 28

As previously we also tested the correspondence between pitch a lightness. Also we used the same idea of design and we used the same squared shape as in the previous correspondence. The result is the following:

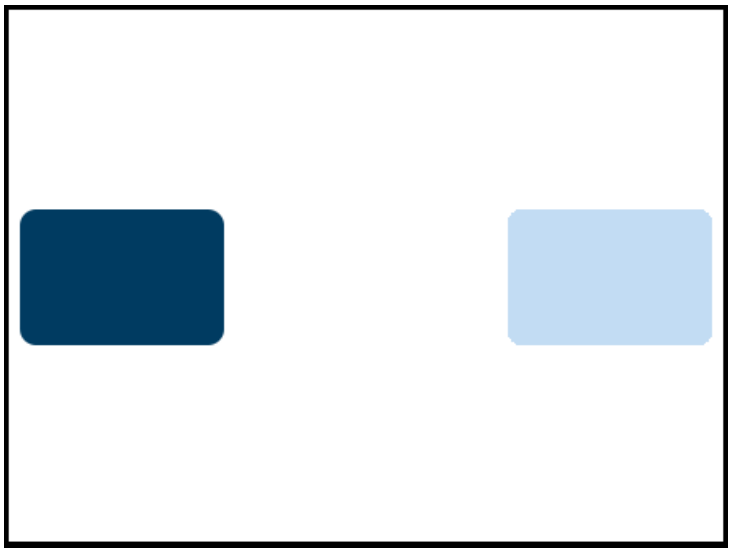

Figure 28

The third part of this group of tests is the one to test with geometric shapes the crossmodal correspondence between pitch (sweep) and spatial position. This one we tested the same way also than before, with a central reference point and two equal shapes situated above and below from the center at the same distance. Figure 29 is the image shown to the participants:

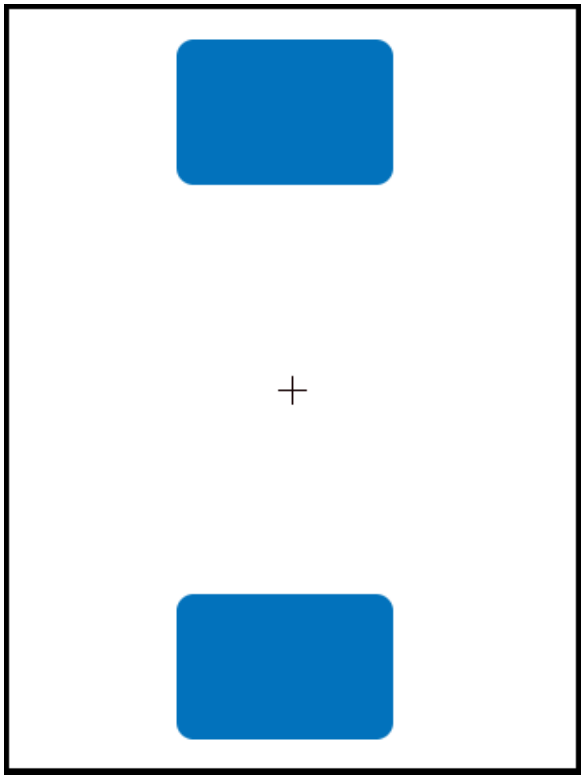

Figure 29

As the last part of the third section of our experiment, with abstract shapes, we tested, as commented previously, the famous metaphoric correspondence of “Takete” and “Maluma”. In this situation we did not make any variation from the actual experiments that are commonly made to test this correspondence. We prepared an image with two shapes really similar to the ones normally used for this experiment, and we shown these ones to the participants. Differently from all the previous tests designed, in this case, due to the nature of the correspondence, we didn’t make a sound play while the kids were staring at the images, but instead we asked the participants which name between “Takete” and “Maluma” would associate to each of the two shapes presented. The image used for this test was the Figure 30.

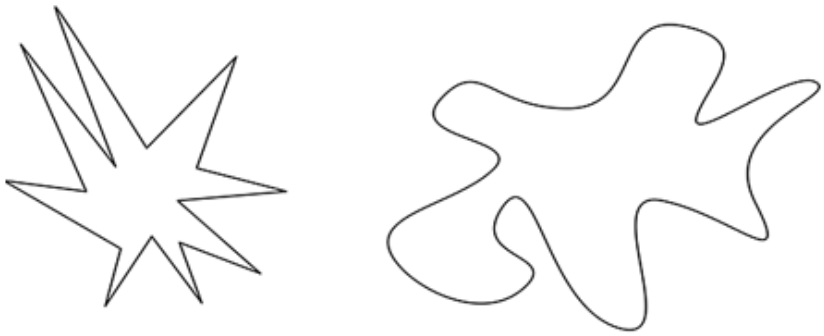

Figure 30. “Takete” - “Maluma” Test

Finally, as the last part of our experiment we decided to test two crossmodal correspondences that we imagined, two crossmodal correspondences way more related to technology, and that we are usually experiencing in our everyday life in a modern society in a western country. Differently and interestingly, in a society where poverty strikes, and the access to technology is lower this correspondence is not that common and makes the experiment have more interest for possible differences.

These two correspondences are the scrolling up/down of a smart device or computer screen and the zooming in/out also in the same platforms. As commented before we

realized that this gestures interacting between the visual movement and the physical, tactile gesture could be really easy and obvious for us but could be different for people that had never been in contact with this technologies or that had a small experience using them.

Therefore in this situation, we did not create a test like the previous ones, because the correspondence was a little bit more complex and was too complicated to test in a similar way than before. In this case we basically let the participants interact with the trackpad of a MacBook Pro of 15 inches, while asking them to proceed doing an action over the tactile pad. First they were asked for scrolling up and down on a document of a kids book, with text and some drawings, and lately to zoom in/out on a picture to get more detail i'm something in particular. And we were analyzing the gestures they were doing intuitively from the very beginning.

4.2.3. Procedure

To make sure the participants were able to complete the experiment and were able to understand every part, and their answers would be valid to be considered to study, we had to follow a procedure. Due to the characteristics and differences between the participants we had to adapt and prepare the tests for their conditions.

Firstly as commented previously the tests had to be prepared to be simple, understandable and attractive for the kids to have interest in participating and getting involved, that is why we had to offer them to the participants as a game with animals, to make them be attracted and excited. That is why we got assessment by an expert in the field, and recommended us to make tests that would be shorter than 3-4 minutes per participant, with a lot of images, and in case the participants were not native english speakers, they would need to have a translator, that would be a familiar person to them, situated close to them but not interfering with the tests that they would be doing.

Following the previous ideas and some others we worked the tests through to get all the information we were needing from the participants. Once the designs were done consciously with the results plotted in the previous section we went to the next steps. We selected, as commented previously, the participants from a Xhosa school with a

range of students from less than 1 year old till 8 years old and the participants from the western school with a range of students from 2 to 12 years old.

The participants, one by one, were explained (the africans non english speakers were getting the explanation translated), before each of them was doing the test, how the process was going to work. They were told that they were going to play a game of association of animal drawings and sounds, and that they had to be really concentrated to do it as well as they could. Firstly they were asked for their name, age, and experience with technology devices, such as smartphones, computers etc. If the participant, was part of the Xhosa group and was having a big experience with technology, was discarded from our experiment, because our main interest was to learn from people with no experience in the field. Once done this first questionnaire, we proceeded to the first test, that was another filter for our participants. Using the Figure 13, the participants were told that they would hear animal sounds and they would have to point on the screen the image of the animal that was making each sound. The participants were having to link correctly at least 3 on 4 to be able to keep doing the test.

Next, the experimenter explained that he was going to show a serie of images, containing two animals of the same type in the same image, on the screen, and the participants should focus on them, while two sounds produced by those two animals would be playing for each image, and once listened both sounds they would need to associate the first animal sound with one of the two animal drawings in the image and the second sound with the other one. In case the participant doubted and needed a second listening, he or she could ask for it and would be played again. This explanation was repeated for every new image that was shown, for the participants to have it in mind and do the activity property. As commented before, the images shown to the participants were shown in different orders to every participant, and using the two examples of each image (counterbalanced versions).

We realized by interacting with the participants and experimenting a little bit, that depending on which of the two age groups were doing the experiment, the amount of tests or activities that they were asked to do, would need to be different, because the youngest group would not be able to stay as long as the older one focused on the

activities and giving interesting results for our study. Therefore , in this second section of the experiment, the participants from the oldest group were viewing and answering a total of 12 images and there were played a total of 24 sounds while the younger kids from the other group were viewing a total of 6 images and were played a total of 12 sounds. After each answer to every experiment, the experimenter was writing down the answers and noting in case there were special situations, like participants not knowing what to answer, or saying that both sounds were produced by the same animal or others.

Once this section was finished, the experimenter explained to the participants that they were going to do the same as done in the previous section but now instead of animals they would be seeing rectangles, and hearing other sounds, but they would have to proceed with the same idea, matching each of the two sounds reproduced with each of the two rectangular shapes that they thought was producing each sound. In this case, both age groups were viewing and listening the same number of images and sounds, 3 images and 6 sounds, also shown aleatory to the different participants.

Following this previous part, the experimenter showed on the full screen of the computer Figure 30 with two shapes, he explained to the participants, while they were watching the image, that those two shapes were having a name; one was called one way "Takete" and the other one another way "Maluma". Then the experimenter asked the participants to focus on the image and he would say two names, and the participants would have to match each name with the shape that they thought was appropriate for that name. The names were repeated by the experimenter two times, before the participants were able to say what they thought, to make sure they were having the actual names on their heads and they were understanding them well.

As the last part of the experiment the experimenter let the participants get hands on the computer. But before that he explained that the track pad of the Mac, was the way of interacting with the screen, and that each move they would do on it, would be seen on the screen. The experimenter showed a few examples to each kid for them to just understand what he was talking about (without doing the gestures that they would be asked followingly). Next he explained he was going to show a book for them on the computer, with a story and some drawings showing the story, and the pages were going to be shown one under the other one, so to go to the next page they would need to go

down on the screen, and they would need to do this while interacting with the trackpad to move the pages, also they were told that the gesture to interact with a page like the one they were going to see is to move up or down two fingers touching the surface. The Experimenter would ask to the participants different type of questions to see how they were interacting with the trackpad trying to get where they were asked. They could be asked: Could you go to the next page? Could you tell me which animal appear on the previous page? Which color is the animal in the next page?

Finally the experimenter explained to the participants that with the same track pad they could make zoom in or out in the image they will have in front. As in the previous exercise, the experimenter, would ask the kids to show him something on the image in a bigger dimension, or to tell him about a detail on that image, prior to that he would show the kids the two possible gestures to do while trying to zoom in or out. In both activities, the experimenter took note of the way the participants were interacting with the trackpad, which gestures they were trying first, which way they trying to get to the point etc.

4.3. Results

As mentioned previously the two age groups were tested the same way, but while the older group answered at the whole collection of items prepared with all the variations (different animals), the younger group was only to two per type of question, that means not answering all the variations prepared. That was done to make the participants of this group be able to answer all the questions more focused and properly and not arrive at the end of the experiment not interested anymore and giving us not valid answers.

In this section we will show the answers of both groups of both cities in form of charts to give an idea of the results and also to be able to compare between both cities and between ages, to understand the results and be able to arrive at some conclusions. Even though the younger group passed less tests than the older one, we weighted the answers to represent them in the same scale as the others to be able to compare them easier. In the charts below the information is represented in a scale from 0 to 100, with two different colors on each bar, each bar represents all the answers from one question from that specific age and culture group. As we can read in the side notes, the darker color

represents the conventional answers while the lighter one represents the non-conventional answers. By conventional we mean those answers that we have been expecting to get in a western country due to the information we got from other previous experiments done by different researchers.

We have to remind that all the results shown in these graphs are the ones we considered valid. As mentioned in previous sections, we discarded data from some participants due to different situations. Firstly are represented the answers from the four different groups to the first six types of questions, testing the three crossmodal correspondences explained before and represented in figures; from 17 to 27

To be able to understand the charts is important to know that the X axis represents the 6 different tests, the three first ones are the ones designed with animals and the next three are the ones made with rectangualar forms. On the Y axis are represented the percentatges of conventional answers (darker shade of grey) obtained form the participants.

4.3.1. Pitch and Size, Lightness and Position Crossmodal Correspondences:

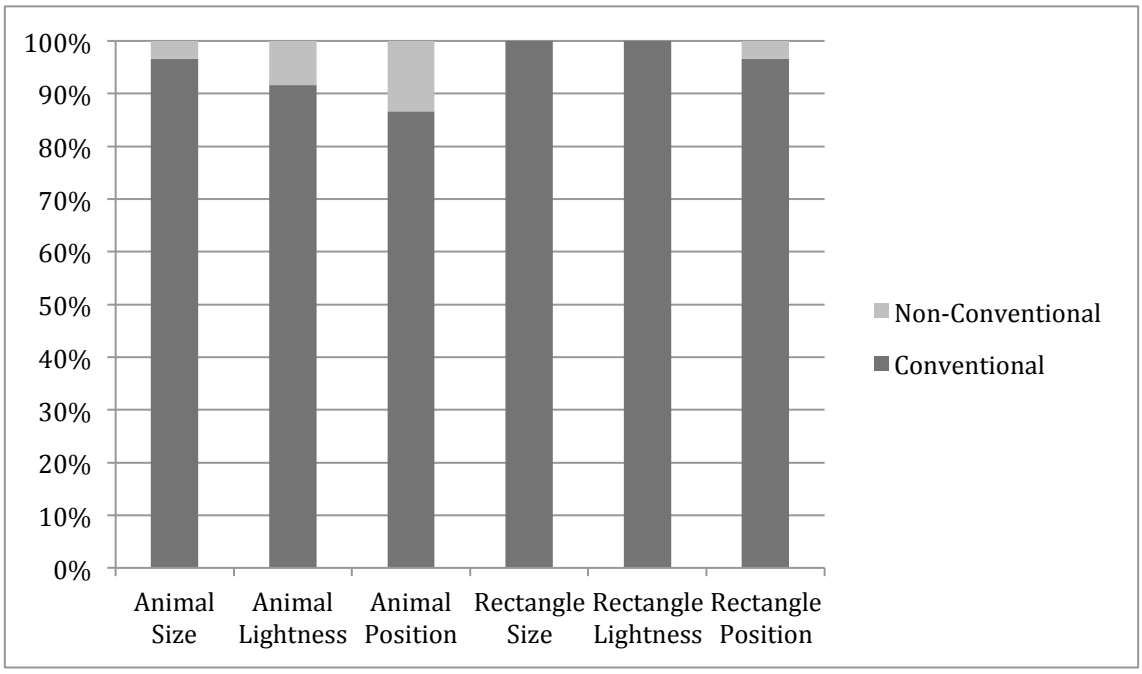

Figure 31. Answers from the Older children From the Xhosa Tribe

In this first representation in figure 31 we can see some interesting characteristics, firstly we see how the three first tests (from Figure 17 to Figure 27) "Animal Size", "Animal Lightness" and "Animal Position" got a high number of conventional answers as well as the three tests using abstract figures, but, we can see how the abstraction step done in the tests; "Rectangle Size", "Rectangle Lightness" and "Rectangle Position", changing animal drawings and animal sounds by rectangle drawings and pure tones made the results be even higher in the conventional side. We can also see how the position test, is the one with lower conventional results in both the animal test and the rectangle test.

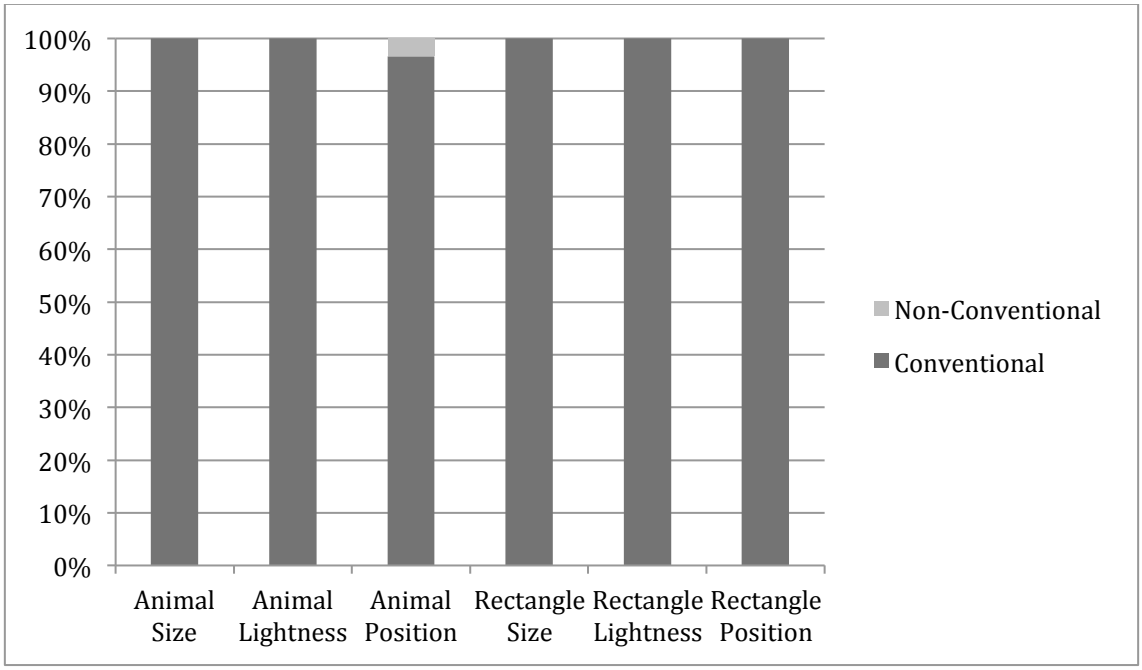

Figure 32. Answers from the Older children From Barcelona

In this chart (Figure 32), representing the answers from the western kids between 6 and 7 years old, we can see how the conventional answers are really high, making in 5 of 6 of the questions have a total of answers as we were expecting in this culture. We can see that the only test that did not get all the conventional answers is the same one that ("Animal Position Test") got the lower conventional answers on the previous chart.

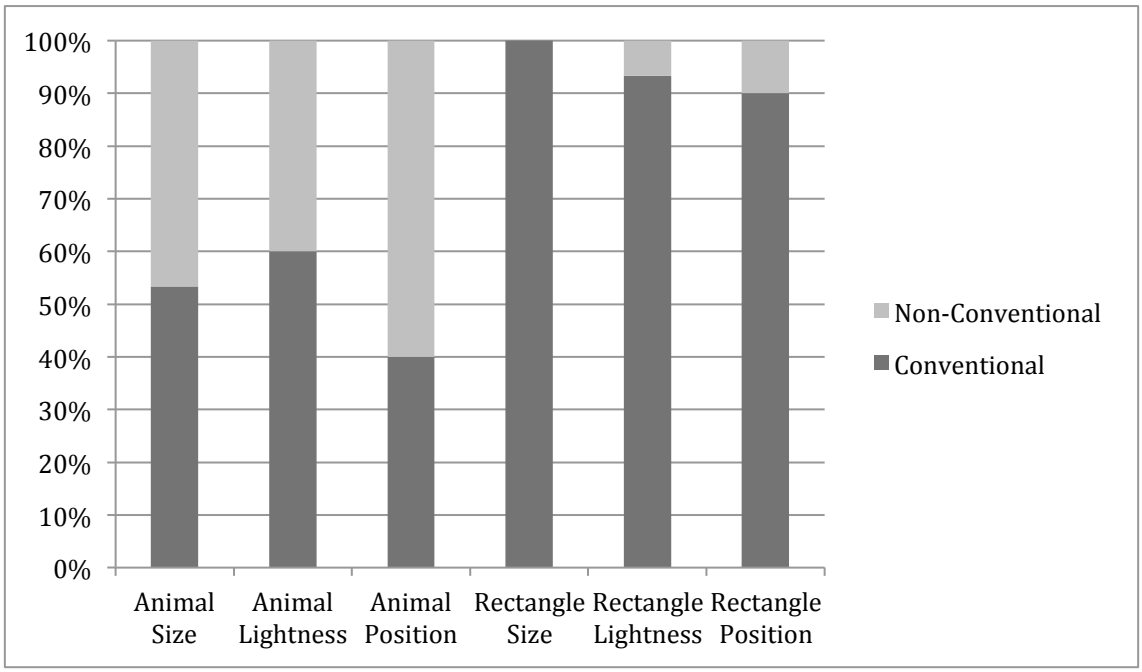

Figure 33. Answers from the Younger children From the Xhosa tribe

In Figure 33 shows the answers given by the younger Xhosa kids, from 3 to 4 years old. We can see easily how in this case the answers are really different from the ones given by older participants from the same cultural group. We can see how the three first questions ("Animal Size" (Figure, 17, 18, 19), "Animal Lightness" (Figure 20, 21, 22) and "Animal Position" (Figure 23, 24, 25)) have a really big number of non-conventional answers. And we can also see how the third question is still the one with the lower rate of conventional answers. Also is clear the difference between the three first questions and the next three, where the last ones get as in the two previous groups a higher conventional rate.

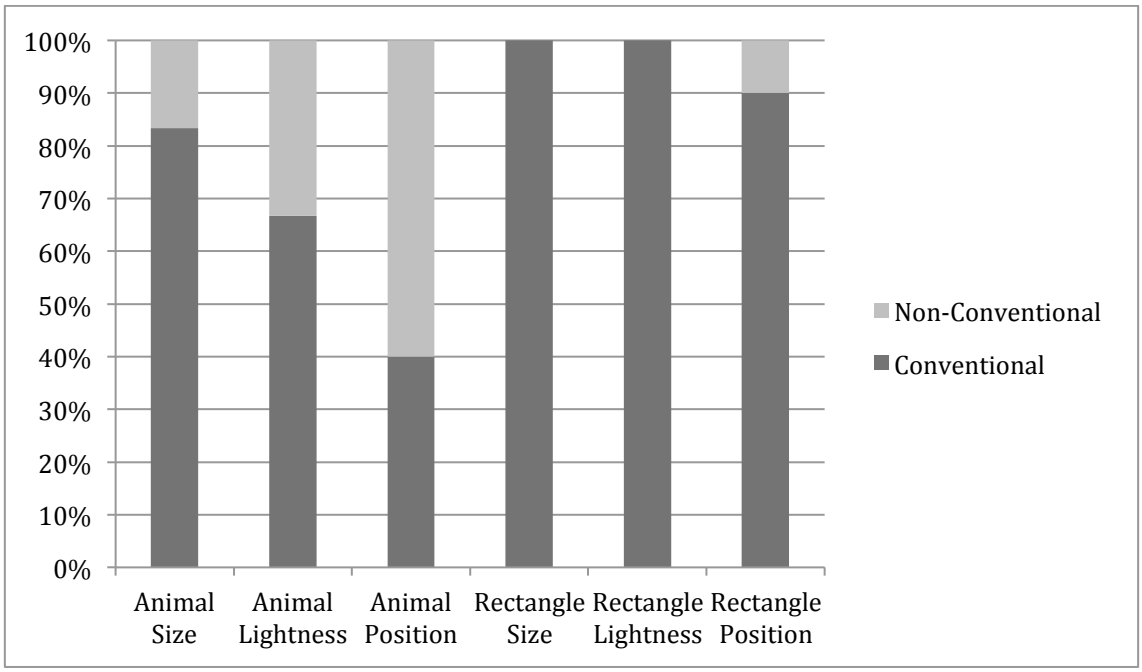

Figure 34. Answers from the Younger children From Barcelona

Finally in Figure 34, we can see how the effect is similar as in the chart of Figure 33. The three first questions have a lower conventional rate than the next three, and also the third question is as in all the other past groups the one with the higher non-conventional number of answers. Also we can tell that in this case the general conventional rate is higher than in the same age group from the Xhosa kids.

4.3.2. Symbolic Correspondence: Takete – Maluma:

Next, as commented in the previous sections we tested the metaphoric crossmodal correspondence "Takete" and "Maluma" represented in Figure 30. The results we got are not needed to be represented in charts, because they are quite homogeneous that makes it easier to be just explained: In the experiment made in South Africa, all the Xhosa kids, independently from their age, answered the conventional answer, that means that they associated "Takete" with the angular shape while "Maluma" was associated to the rounded shape. Differently and surprisingly, in the test made in Barcelona to the western kids, both age groups associated non-conventionally the nouns and the shapes, so they associated "Takete" with the rounded shape, while "Maluma"

with the angular one, but when repeating the test with different words like “Kiki” and “Maluma” the results went back to the conventional ones. Later on we provide an interpretation of this result, based on linguistic factors.

4.3.3. Technological Crossmodal Correspondences:

Finally, following the order explained in the previous section, the kids from both cultural groups were asked to scroll up and down a digital kids book on the computer and also to zoom in and out an image on the same surface. The answers obtained for the scrolling up and down question were the following: of the group of younger kids from the Xhosa tribe 7 out of 15 moved the two fingers up to find something in pages below they were in. In the older group of the same cultural side, 5 out of 15 moved the fingers up. In the younger group from the western kids, 3 out of 15 also moved the fingers up to find something situated below, and in the older group from the kids from Barcelona, only 1 out of 15 did the same move just described. We can see this information represented in Figure 35, where on the X axis, we find the four different groups tested, and on the Y axis we find their number of answers. The question in all of them was to find something in below pages, and their answer is represented in a darker shade of grey if they scrolled their fingers up to find something below where they were or in a lighter shade of gray if opposite.

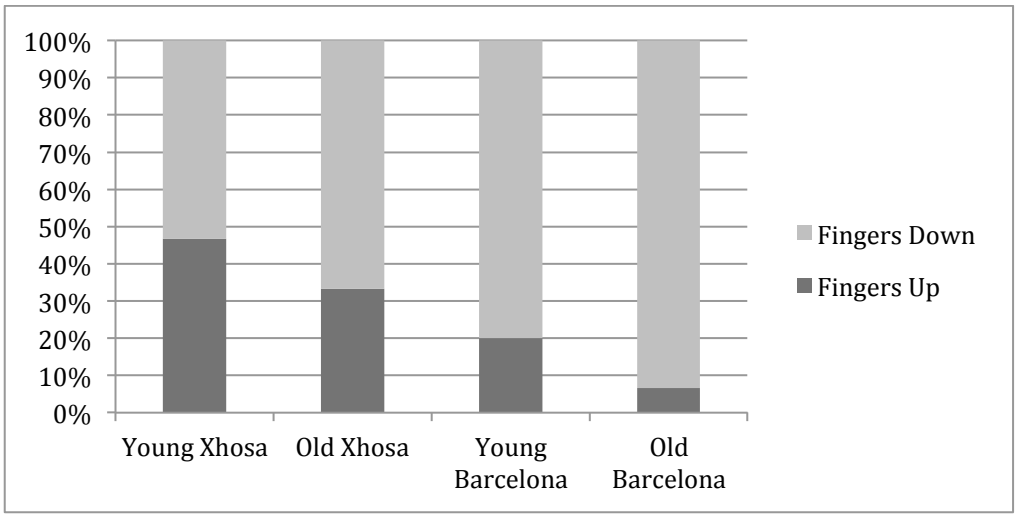

Figure 35. Answers From Scroll Up-Down Test

The answers obtained for the zooming in and out test were the following: in the older group of the Xhosa community as well as younger and older group of the western kids, the totality of the participants zoomed in the conventional way, moving the fingers from separated to together, while in the younger group from the Xhosa tribe, 3 out of 15 of the participants did it the opposite way. The results are represented in Figure 36, where to the question of zooming into something, the darker shade of grey representes the participants that joined their fingers, while in the lighter shade the ones that separated them:

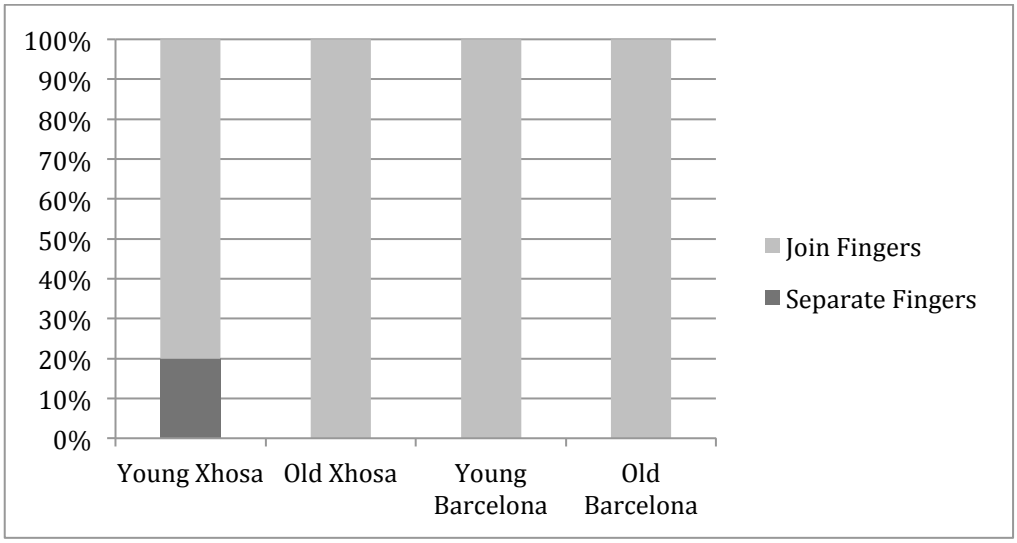

Figure 36. Answers From Zoom In-Out Test

4.4. Results Interpretation

4.4.1. Pitch and Size, Lightness and Position Crossmodal Correspondences:

While studying the different results we obtained from the experiment there are some characteristics that are easy to interpret and and some other aspects are more complicated and require a deeper study. We can see easily how some questions repeated in the different tests have the same answers over the different cultural and age groups. We can tell for example by looking at the charts, how the pitch – position ("Animal Position") question, in three of the four groups has the lowest rate of conventialonal answers of all the different questions. Those answers come from the test of the crossmodal correspondence between pitch and spatial position.

Even if the results in most of the answers are not too low, we think that the results obtained from the different questions of "Animal Position" could be interpreted in two different ways: They could mean that the design of the test wouldn't be appropriate, or at least proper for this specific participants and that they would not comprehend well the system or see the connection between the pitch sweep and the images. But also could be that this type of crossmodal correspondence, being part of the "linguistic correspondences" would not arise in children as young as these ones, because they need a bigger perfection of the vocabulary of the language to associate the words high or low with auditive stimulus and with the spatial position. This last interpretation would be also based on the fact that the younger the group tested, the more non-conventional the answers from this test, so that could mean that while getting older the experience plays a big role in the absorption or the perfection of this crossmodal correspondence.

Another aspect we can appreciate by looking at the results is the fact that the answers given by the two younger groups, in both cases are lower in conventional number than the answers obtained from the older kids, that also could reflect the fact that experience plays a big role in the crossmodal correspondences. Yet, here we realize of what we think are some of the most interesting aspect that these results offer. By observing the results of the two younger groups, we can see a big difference on the answers obtained while testing them with the animal designs or with the abstract rectangular designs. The answers obtained from the animal tests, are quite low in conventional answers if we compare them with the older groups, but also if we compare them with the same crossmodal correspondence tests but with the abstract designs, we can see how the differences are really big.

This tells us how the younger kids could have been distracted by interacting with the animal designs, having other concepts in their head that influenced more on them than the actual crossmodal correspondence between the sound and the image, but when the concept was presented in abstract form, they did not have those distractions and we can see how the results are really high, almost hundred per cent conventional in all the tests. That also makes us think, that even if experience could be really important in this phenomenon of the crossmodal correspondences, it seems that there could be a part of it innate.

The fact that the participants chosen were not the youngest they could be, due to many different situations, does not let us investigate forward in this matter, and does not let us compare results from these ages already tested and newborn children that would maybe make things be clearer.

4.4.2. Symbolic Correspondence: Takete – Maluma:

Following the order of our experiment, the next test we did was the famous "Takete" and "Maluma" and as commented in the previous section, all of the answers were conventional, except the tests made in Barcelona. But this situation has it's reason. All the participants from the western group, explained that they chose to link "Takete" and the rounded shape because the word "Takete" was sounding like "taqueta" that means little stain in Catalan, their mother tongue language. That is why we decided to redo the rest using "Kiki" instead of "Takete" and then the results became conventional again.

4.4.3. Technological Crossmodal Correspondences:

Finally from the results obtained from the technological correspondences that we decided to test, we got as mentioned before a majority of participants reacting the conventional way, every group got a higher rate of conventional interactions, but the Xhosa tribe participants had a lower number than the western participants. That is due to the lack of experience that they had with this devices and the big difference with the children from Barcelona that had a big experience on the field, this is really palpable in Figure 35 and Figure 36. Also by comparing the results between both Xhosa groups, we can see how the younger ones have a lower rate on their answers on conventional responses, while both age groups have the same lack of experience, that could mean that the younger children didn't undertand well the instructions. By regarding the globality of the answers we can say that these crossmodal correspondences and conventional way of interacting with technological devices are the most natural way of doing it, for both cultural groups.

5. POSSIBLE FUTURE PROTOTYPE

As explained along this document, one of the objectives of this project is to obtain conclusions from the experiments done and create a list of ideas that could help in future technological designs. From the interpretations of the results obtained in the last section and some ideas we got while making the tests, we can list a few ideas we think could be interesting and useful for future designs to help the user have a better interaction and a more pleasant feeling.

One of our hypothesis was that we could maybe find some relevan differences between cultures, while experimenting in a western city like Barcelona and Mitchells Plain, Western Cape, South Africa. After the experiment was done, we observed that the differences between both cultures, in the specific tests and crossmodal correspondences that we chose, were not big, to be considered as something really relevant. That information gave us the conclusión that the posible idea we had at the beggining of working in the future on cultural centered design makes no sense reviewing the results, so it would be more appropiate to adapt a common design for both cultures, due that the differences are not really visible in the experiment we tested.

Therefore, for the general design, we obtained relevant information from our results that prove one of the biggest ideas we had in mind, the fact that crossmodal correspondences used propertly can guide the attention of the users. In this case between vision and audition we proved how the variation of a pitch was affection the where the participants were focusing their vision. This idea let us the clear possibility of using it for a future better intuitive and guided design, to make interaction easier and better.

Let’s imagine that somebody is interacting with a smart device, and this person needs to click something on the screen to go to the next page, or needs to find an object over many others. Then, what if by adding a congruent sound with what this person is trying to find helps the action to be more intuitive and easy. Obviously, also, regarding the results, we would have a bigger success using the correspondences that got a higher rate in conventional answers. So for example, we could use a high pitch to guide the vision to the small object the user is trying to locate, or the lighter one, or the one located

higher on the screen. We could use these concepts to create pop-up with the rest of distractors from the screen and guide the visual attention.

We think that this concept is really interesting and useful and that different combinations of it could generate a really intuitive interaction that would make the user feel good while interacting and would create a better sensation and appreciation of the interactive and it's content.

Regarding the other tests done, we observed how the technological correspondences got a majority of conventional answers in both cultures, so that makes us think that in a future design we would also use this two interactive gestures the same way they are nowadays, because we considere them to be the most natural and intuitive ways to opperate with the system.

Finally there is an important aspect we observed while studying the answers from our Pitch and Size, Lightness and Position tests, that we consider really relevant and interesting to have in mind for future design. The fact that both young groups in both cultures got a visible difference of conventional responses between when the questions were made with the animal shapes and the abstract shapes, when in the animal shapes they were not percieving that easy the connection between the two sences, and when the rectangular shapes they were giving almost 100% conventional answers, associating each pitch with each size, lightness or position, and the fact that this answers are different from the older groups, makes us think that the general design should probably be adapted or at least having in mind the different ages of the potential users to make them interact propertly. We think that for younger users, even if trying to make a more attractive design by using aniaml drawins or others, is important to keep distractions away if what we want is creating a useful and intuitive design that Works.

6. CONCLUSIONS AND FUTURE WORK

After all the project has been done and we have seen the results from the experiment generated, we got some conclusions and some ideas for the future, to make this piece of work be usefull in future projects or to be extended.

From the first objectives we had in mind while starting this project, we can say that we obtained the results that were expected in most of the cases and we got enough information and knowledge for giving the ideas or conceps of future prototypes that we explained previously, that is why we consider that our project has acomplished our expectations. Never the less, some aspects of the project and some ideas we had at the beggining did not turn the way we thought, but that was part of the main idea of this experiment that we decided to do.

We did this work and tested the experiment in specific conditions, in a defined period of time and with a low experinece on this field, but with hard work and help from experts in the matter, we ended creating something that we consider really interesting regarding our possibilities. That is why we think that this project could be a start of a longer idea or work, that with more time and better conditions could become even more interesting and cover a bigger field.

We think that many aspects of our project could be varied on the future or extended. We think that this idea of experiment is extrapolable to many other cultures not studied yet and also extended to many other age ranges, to obtain more information for a better design for everyone, as that is our main goal. If we consider our results obtained from our different age groups, we can remind that one of the biggest conclusions of those answers, is the fact that the younger groups of students seem to not respond to the crossmodal correspondences while answering the “animal” tests, but these same participants answer conventionally when asked the same questions with abstract shapes, with no distractions, that shows u show this crossmodal correspondences that seemed to not be part of these participants are actually “inside” of them since they are really young. We think that if instead of have tested the differences between age groups of 6-7 years old and 3-4 we would have changed the younger group for an even younger one, preferibly new-born children, we would have obtained more relevant results that would

have said more of the actual differences and of the innate – learned discussion of the crossmodal correspodences. That is why we think that this could be a really interesting way of continuing our project, testing new-born children to figure out the answers from this dicussion of innate- learned. That would mean, that it would be needed a different type of technology for testing this individuals and a more specific design of the experiment to adapt them to this group, and would add complexity to the future project.

Also while working on this project we had to choose a finite number of correspondences to test, the ones we though more convenient and that would be easier for us to experiment with the specific groups we chose. After obtaining the different results from our experiment, we think that in the future could be really interesting to test other correspondences that could extend the knowledge on the field. These correspondences could be tested in the same cultural groups we did, to see differences, and maybe in some of this could be posible to obtain bigger differences in the results between both cultural groups.

We have tested mainly correspondences between vision and audition, because as commented, were easier for us to test with what we had, and we thought would be some of the most interesting ones. But we think that with the propper technology and planning could be really interesting to follow the same idea to test correspondences between other senses, mainly between audition and taste. Because we think that this two senses could show a lot of differences between cultures and could be really used for future designs, or even for more marketing consumer aspects. Lately the interest for consumer psychology has been increasing and we think that this cultural study could be really interesting and informative in this matter, and would help understand how to design all types of products, from packaging to technological, interactive devices.

7. BIBLIOGRAPHY

7.1. Papers

- Bernstein, I. H., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, *87*(2), 241.
- Bond, B., & Stevens, S. S. (1969). Cross-modality matching of brightness to loudness by 5-year-olds. *Perception & Psychophysics*, *6*(6), 337-339.
- Braaten, R. (1993). Synesthetic correspondence between visual location and auditory pitch in infants.
- Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). "Bouba" and "Kiki" in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*(2), 165-172.
- Cabrera, D., & Morimoto, M. (2007). Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs. *The Journal of the Acoustical Society of America*, *122*(1), 478-488.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral cortex*, *11*(12), 1110-1123.
- Cyr, D., & Trevor-Smith, H. (2004). Localization of Web design: An empirical comparison of German, Japanese, and United States Web site characteristics. *Journal of the American society for information science and technology*, *55*(13), 1199-1208.
- Davis, R. (1961). THE FITNESS OF NAMES TO DRAWINGS. A CROSS-CULTURAL STUDY IN TANGANYIKA. *British Journal of Psychology*, *52*(3), 259-268.
- Deroy, O., & Auvray, M. (2013). A new Molyneux's problem: Sounds, shapes and arbitrary crossmodal correspondences. In O. Kutz, M. Bhatt, S. Borgo, & P. Santos (Eds.), *SHAPES* (pp. 61-70).
- Deroy, O., & Valentin, D. (2011). Tasting liquid shapes: investigating the sensory basis of cross-modal correspondences. *Chemosensory Perception*, *4*(3), 80-90.
- Diffloth, G. (1994). I: big, a: small. *Sound symbolism*, 107-113.
- Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision*, *10*(1), 6.

- Fernández-Prieto, I., Navarra, J., & Pons, F. (2015). How big is this sound? Crossmodal association between pitch and size in infants. *Infant Behavior and Development*, *38*, 77-81.
- Fontanari, L., Gonzalez, M., Vallortigara, G., & Girotto, V. (2014). Probabilistic cognition in two indigenous Mayan groups. *Proceedings of the National Academy of Sciences*, *111*(48), 17075-17080.
- Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics*, *68*(7), 1191-1203.
- Grossenbacher, P. G., & Lovelace, C. T. (2001). Mechanisms of synesthesia: cognitive and physiological constraints. *Trends in cognitive sciences*, *5*(1), 36-41.
- Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology*, *8*(6), R190-R191.
- Knöferle, K., & Spence, C. (2012). Crossmodal correspondences between sounds and tastes. *Psychonomic bulletin & review*, 1-15.
- Köhler, W. (1929). Gestalt psychology.
- Köhler, W. (1947). Gestalt psychology: The definitive statement of the Gestalt theory. *Liveright, New York*.
- Lalanne, C., & Lorenceau, J. (2004). Crossmodal integration for perception and action. *Journal of Physiology-Paris*, *98*(1), 265-279.
- Leiber, P., & Spanner-Ulmer, B. Cultural differences in the use of mobile devices.
- Lewkowicz, D. J., & Turkewitz, G. (1980). Cross-modal equivalence in early infancy: Auditory–visual intensity matching. *Developmental psychology*, *16*(6), 597.
- Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, *108*(51), 20661-20665.
- Madden, T. J., Hewett, K., & Roth, M. S. (2000). Managing images in different cultures: A cross-national study of color meanings and preferences. *Journal of international marketing*, *8*(4), 90-107.
- Marks, L. E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *The American journal of psychology*, 173-188.
- Marks, L. E. (1987a). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 384.

- Marks, L. E. (1987b). On cross-modal similarity: Perceiving temporal patterns by hearing, touch, and vision. *Perception & psychophysics*, *42*(3), 250-256.
- Marks, L. E. (1989a). For hedgehogs and foxes: Individual differences in the perception of cross-modal similarity. In *Psychophysics in action* (pp. 55-65). Springer Berlin Heidelberg.
- Marks, L. E. (1989b). On cross-modal similarity: the perceptual structure of pitch, loudness, and brightness. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 586.
- Martino, G., & Marks, L. E. (2000). Cross-modal interaction between vision and touch: the role of synesthetic correspondence. *PERCEPTION-LONDON-*, *29*(6), 745-754.
- Martino, G., & Marks, L. E. (2001). Synesthesia: Strong and weak. *Current Directions in Psychological Science*, *10*(2), 61-65.
- Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 69.
- Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 133-136.
- Nardo, D., Santangelo, V., & Macaluso, E. (2014). Spatial orienting in complex audiovisual environments. *Human brain mapping*, *35*(4), 1597-1614.
- Ngo, M. K., & Spence, C. (2011). Assessing the shapes and speech sounds that consumers associate with different kinds of chocolate. *Journal of Sensory Studies*, *26*(6), 421-428.
- Ngo, M. K., Piqueras-Fiszman, B., & Spence, C. (2012). On the colour and shape of still and sparkling water: Insights from online and laboratory-based testing. *Food Quality and Preference*, *24*(2), 260-268.
- Noiwan, J., & Norcio, A. F. (2006). Cultural differences on attention and perceived usability: Investigating color combinations of animated graphics. *International journal of Human-computer studies*, *64*(2), 103-122.
- Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic books.
- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of experimental child psychology*, *114*(2), 173-186.

- Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound–shape cross-modal correspondences in 4-month-olds. *Journal of experimental child psychology*, *114*(2), 173-186.
- Parise, C. V. Crossmodal Correspondences: Standing Issues and Experimental Guidelines.
- Parise, C. V., & Spence, C. (2009). When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One*, *4*(5), e5664.
- Parise, C., & Spence, C. (2008). Synesthetic congruency modulates the temporal ventriloquism effect. *Neuroscience Letters*, *442*(3), 257-261.
- Parise, C., & Spence, C. (2013). Audiovisual cross-modal correspondences in the general population. *The Oxford Handbook of Synesthesia,*, 790-815.
- Peña, M., Mehler, J., & Nespor, M. (2011). The role of audiovisual processing in early conceptual development. *Psychological science*, *22*(11), 1419-1421.
- Persson, V. (2011). Crossmodal correspondences between visual, olfactory and auditory information.
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological science*, *24*(6), 1037-1043.
- Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499-503.
- Piqueras-Fiszman, B., & Spence, C. (2011). Crossmodal correspondences in product packaging. Assessing color–flavor correspondences for potato chips (crisps). *Appetite*, *57*(3), 753-757.
- Roffler, S. K., & Butler, R. A. (1968). Factors that influence the localization of sound in the vertical plane. *The Journal of the Acoustical Society of America*, *43*(6), 1255-1259.
- Root, R. T., & Ross, S. (1965). Further validation of subjective scales for loudness and brightness by means of cross-modality matching. *The American journal of psychology*, 285-289.
- Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition*, *99*(2), 113-129.
- Sagiv, N., & Ward, J. (2006). Crossmodal interactions: lessons from synesthesia. *Progress in brain research*, *155*, 259-271.

- Sapir, E. (1929). A study in phonetic symbolism. *Journal of experimental psychology*, *12*(3), 225.
- Simpson, R. H., Quinn, M., & Ausubel, D. P. (1956). Synesthesia in children: Association of colors with pure tone frequencies. *The Journal of genetic psychology*, *89*(1), 95-103.
- Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971-995.
- Spence, C., & Deroy, O. (2012). Crossmodal correspondences: Innate or learned? *I-Perception*, *3*(5), 316–318. doi:10.1068/i0526ic
- Spence, C., & Deroy, O. (2014). Tasting shapes: a review of four hypotheses. *Theoria et Historia Scientiarum*, *10*, 207-238.
- Spence, C., & Gallace, A. (2011). Tasting shapes and words. *Food Quality and Preference*, *22*(3), 290-295.
- Spence, C., & Ngo, M. K. (2012). Assessing the shape symbolism of the taste, flavour, and texture of foods and beverages. *Flavour*, *1*(1), 1-13.
- Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology*, *13*(13), R519-R521.
- Spence, C., Levitan, C. A., Shankar, M. U., & Zampini, M. (2010). Does food color influence taste and flavor perception in humans?. *Chemosensory Perception*, *3*(1), 68-84.
- Stevens, J. C., & Marks, L. E. (1965). Cross-modality matching of brightness and loudness. *Proceedings of the National Academy of Sciences of the United States of America*, *54*(2), 407.
- Vyncke, F., & Brengman, M. (2010). Are culturally congruent websites more effective? An overview of a decade of empirical evidence. *Journal of Electronic Commerce Research*, *11*(1), 14-29.
- Wagner, K., & Dobkins, K. (2009). Shape-color synesthesia in the first year of life: A normal stage of visual development?. *Journal of Vision*, *9*(8), 699-699.
- Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). " Metaphorical" Mapping in Human Infants. *Child Development*, 728-731.
- Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., & Johnson, S. P. (2009). Preverbal Infants' Sensitivity to Synaesthetic Cross-Modality Correspondences. *Psychological Science*.

- Wan, X., Woods, A. T., van den Bosch, J. J., McKenzie, K. J., Velasco, C., & Spence, C. (2014). Cross-cultural differences in crossmodal correspondences between basic tastes and visual features. *Frontiers in psychology*, *5*.
- Wicker, F. W. (1968). Mapping the intersensory regions of perceptual space. *The American journal of psychology*, 178-188.

7.2. Websites

- Interaction DesignFoundation
[https://www.interaction-design.org]
Consulted the 20/01/2015

- South African Information Website
[http://www.southafrica.info/about/people/language.htm#english]
Consulted the 15/04/2015

- Mitchells Plain Wikipedia Page
[https://en.wikipedia.org/wiki/Mitchells_Plain]
Consulted the 16/04/2015

- Languages of South Africa Wikipedia Page
[https://en.wikipedia.org/wiki/Languages_of_South_Africa]
Consulted the 17/04/2015

- [1] Gestures Interactive Design
[https://en.wikipedia.org/wiki/File:Gestures_Two_Finger_Scroll.png]
Consulted the 22/05/2015

- Participatory Design Wikipedia Page
[https://en.wikipedia.org/wiki/Participatory_design]
Consulted the 12/06/2015

8. APPENDIX

We list some information in the following lines that could be interesting for some readers to understand better the cultural situation in the Western Cape, South Africa:

Afrikaans: Is the third most common language in the country, spoken by 13.5% of the inhabitants. Afrikaans is spoken mainly by coloured and white South Africans. The origins of the language come from the 17th century Dutch, with some influences from other languages like English, Malaysian, German, Portuguese, French and some other African languages.

Initially called Cape Dutch, this language has most of it's speakers in the Western Cape, where is the mother tongue of almost half of the population of the region. Back in the days Afrikaans was a largely spoken language but with proper "formal" Dutch when written. Also is said that Afrikaans is the Dutch kitchen language, that the white families service people (mates, cleaners, cooks…) created by changing the Dutch in a way that "the Boss" would not be able to understand.

Linguistic lineage: Indo-European → Germanic → West Germanic → Low Franconian → Afrikaans

English: This language has been a really important and influential language in South Africa, adapting itself to the different communities and regions of the country. Around half of the population of the country have a speaking knowledge of the language. English has been an official language in the Cape Colony since 1822 and with the union of the different areas and colonies of South Africa in 1910 it became together with Dutch the official language of the country. And at this time is the language spoken for government aspects, business and commerce.

English is spoken widely in the country by almost 5 million people, some of them are the Indian-South African and Chinese-South African. This language in South Africa has been influenced by Afrikaans and many other African languages that makes it become an special dialect, where many words used are not understandable by other English speakers.

The most English spoken region of South Africa is Guateng where we find a 32.8% of all the English speakers in the country. In the province that we are more interested, Western Cape, English is spoken by 19.7% of the population of the region.

Linguistic lineage: Indo-European → Germanic → West Germanic → English

Xhosa: This is the second most spoken language in the country, with a 16% of all the population, or in other words 8 million people. It is a regional language and it's to biggest spoken regions are the Eastern Cape with almost 78% speakers from that province and the Western Cape where a 24% of the regional population speak the language, making a 17% of the total of isiXhosa speakers in the whole country. Also in the other regions there are percentages of Xhosa speakers, but not as wide as in the two listed above.

Xhosa is a Nguni language, therefore is a tonal language. It shared a lot of words and grammar with Zulu, due to its origin. This language is known worldwide by the originality of it and the big difference from the languages normally know in the western countries. Is known as the language of clicks, because while speaking there are a number of clicks that are done with the mouth while pronouncing the rest of the word.

Linguistic lineage: Niger-Congo → Atlantic-Congo → Volta- Congo → Benue-Congo → Bantoid → Southern → Narrow Bantu → Central → S group → Nguni → Xhosa

With this definitions we can have a better idea of what is the situation in the region we focused our work. As we can see, in the Western Cape, with a total area of 129,462 km^2, we find three major languages, spoken widely in the whole region, due to the history of it, the influence between past languages in the area and the immigration in the past years